

How and when to halt SDI

A calculation in this issue shows X-ray lasers might be feasible, and points to horrendous difficulties ahead. But the case for putting and end to the search for ballistic missile defences rests elsewhere.

Is it good news, or bad news, that the prospects for using X-ray lasers for destroying other people's strategic missiles are more real than generally supposed? The conclusion is the burden of the important article on page 487 of this issue; for most people, the answer will depend on prejudice. Using published information and ingenious argument, Professor D.M. Ritson of Stanford University has been able to show that a 120 kT fission bomb exploded near a bundle of accurately shaped and oriented zinc wires could loose off pencil beams of laser X-rays at targets such as rockets, at least some of them at 2,000 km. Does that mean a new weapon round the corner? Physically, the effect is simple enough. High-frequency electromagnetic radiation from a fission bomb will normally be degraded, by successive absorption and emission from atoms of the surrounding medium, into black-body radiation corresponding to the temperature of the fireball. What now emerges is that, by physically intercepting this process with a material capable of functioning as an X-ray laser during the few microseconds of its geometrical integrity, it is feasible to keep a larger proportion of the radiation in the form of X-rays and to arrange that this radiation is directed in predetermined directions, not spread uniformly over the surface of a sphere. But Ritson's argument also shows that many more expensive and technically demanding experiments will be required before the system can be proved, let alone installed.

What does that imply for ordinary mortals? This is where the prejudice comes in. Supporters of the Strategic Defence Initiative (SDI) reading Ritson's article will say "We told you so!". And if fission-pumped X-ray lasers can function is not the concept of SDI as a whole attainable? Sceptics, with as much justice, will favour a different reading, noting that the feasibility of the parts does not entail that the whole would function with the precision and the completeness necessary to ensure immunity from attack. They will also observe that the cost of the system would almost certainly be ruinous. This is why the chief consequence of Ritson's argument for the continuing controversy about SDI is not to resolve it, but, by making the argument more realistic, to sharpen it. That is an important development when the Soviet Union has once again made clear (last week) its consistent view that SDI is an impediment to further agreements on the limitation of strategic missiles.

Curious

A curious situation is developing. The Soviet Union and the United States, in their negotiations at Geneva, appear to be moving steadily towards the most radical of agreements on missiles of intermediate range — total abolition. With last week's concession by the United States that Pershing II missiles removed from Europe would not be converted to other uses, there appears to be no obstacle to an agreement except the willingness of those concerned to sign it. Reaching this position, the Soviet Union appears to have accepted that British and French nuclear forces, although technically comparable with those now to be removed, will not (for the time being) impede agreement, and also that there must be limitations on short-range weapons in central Europe as well. Yet the question of strategic arms remains up in the air. The Soviet Union says it would prefer total abolition, but has nevertheless presented the

negotiators at Geneva with a draft treaty allowing for a 50 per cent reduction, with agreement conditional on the abandonment of SDI. Last week, there were ominous hints that even the proposed agreement on intermediate missiles might be linked with the same precondition.

So, after all the hard work since last November's abortive summit at Reykjavik, is the promise of a substantial arms control agreement to be put on ice until a new president is elected in the United States? Or simply allowed to go to waste? Everybody, the Soviet Union included, should know what accomplishments would be lost. European governments, whose anxieties about the deployment of Soviet SS-20 missiles led, in 1979, to the decision that offsetting US missiles should be based in Europe, but which became used to the sense of extra security thus provided, have been persuaded that they cannot be made worse off by a return to previous arrangements. The Soviet Union, similarly, appears to have accepted that an agreement on intermediate missiles now would be worthwhile in itself but also as a necessary prelude to a strategic agreement; that is why it would be folly to insist on a last-ditch link with SDI. The benefits for the United States are both real and symbolic: limiting the military commitment to Europe would be a financial and political plus, while the soon-outgoing administration needs something to boast about (and to help in next year's election).

Commitment

That is why the issue of SDI should be postponed to the next round of negotiations. President Reagan's personal commitment to the enterprise will ensure that. His wish to see an arms control agreement in the history books against his name is probably not as strong as his belief that an effective defence against ballistic missiles would be an historic benefit. So it might be if one could be constructed without risking trouble, in the form of countervailing strategic reinforcement or even pre-emptive strikes, and if there were also arrangements to ensure that none of the nuclear powers used missile defences to be able to attack others with impunity. All this leads to the paradoxical conclusion that SDI, designed to make hostile strategic missiles irrelevant, requires an agreement for the drastic curtailment of strategic weapons before it can be realized.

Given secrecy, and the unavoidable uncertainties (which cut both ways), the outcomes are never clear-cut. The panel of the American Physical Society that spent two years studying the question, and whose report will be published in *Reviews of Modern Physics* later this month (see *Nature* 327, 263; 1987), appears to have run into trouble from those protesting that its list of formidable obstacles to be overcome has been used for political purposes (and that errors have survived the refereeing process), but that is hardly the point. Supporters of SDI claim that the outstanding problems must be solvable, but that will be possible only at huge cost and by the testing of system components in ways that must violate the ABM treaty. That is why attempts to calculate the performance of SDI are in a sense irrelevant; what matters more is that they illustrate the enormous investment in research there will have to be to bring the project within reach of viability. Would it not be simpler just to call a halt with an agreement between the United States and



the Soviet Union never to deploy in space any piece of equipment capable of destroying another (weapons of "mass destruction" are already banned)? The deployment of SDI components would be one of those rare and recognizably novel occasions in military affairs when something qualitatively novel is being attempted. Those who look for firebreaks in the arms race would be right to think that this is one. If, in the short run, the Soviet Union is compelled to regard President Reagan's attachment to SDI as a fact of life, may not Mr Mikhail Gorbachev, in the next round, require the United States to take his determination that SDI shall not prosper in the same spirit? □

Clandestine government

The US government needs an effective cabinet; British government should not gag its press.

THERE is an odd parallel between the congressional investigation in the United States of the sale of arms to Iran (and the diversion of some of the funds collected) and the present fuss in Britain about the memoirs of Mr Peter Wright, a former agent of the British security services now living in Tasmania. But there are also glaring contrasts between the two affairs. The proceedings in Washington are remarkable for how much has been told of how public servants set about running clandestine operations of which the government which employs them was collectively ignorant.

In the British case, on the other hand, where one of Wright's revelations (in a book called *Spycatcher*) is that the intelligence services had set out in the 1970s to discredit those to whom it was responsible, the government has gone to extraordinary lengths to make sure that Wright's allegations, true or false, are not discussed. Last week, the government won a ruling from the legal committee of the House of Lords that the British press may not discuss the details of the allegations which are probably being given more publicity elsewhere than they deserve by the government's attempts to keep them quiet.

The lesson to be drawn from what has emerged in Washington is that the United States needs a device for making cabinet government function as it should. The tale is that a handful of people employed at the White House were able to carry out a clandestine operation without formal authority and without most members of the government knowing what was going on.

What the congressional committee has failed to do is to discover whether President Reagan knew what had been going on. None of the witnesses has implicated the president, but most have blamed the late director of the Central Intelligence Agency, William Casey, who died a few days after first giving private evidence to the committee, and who was a member of the cabinet. Two other cabinet members (the Secretaries of State and of Defense) said that what little they learned of the clandestine operation led them to protest, with the consequence that they were shut out from further discussions of the matter. That could not happen in Britain, where any member of the cabinet has a right to know what may be going on, and in return is required to support the collective view even if he does not like it — or to resign. On the second centenary of an otherwise admirable constitution, the congressional committee should look for ways of asserting collective cabinet responsibility in the United States.

In Britain, the doctrine works, if anything, too well. There is no reason to believe that the twenty-odd members of the cabinet all believe that *de facto* censorship of the British press on a particular matter is an appropriate price to pay for the pursuit of legal means for preventing discussions of the Wright book, especially because the effort will eventually be fruitless. The government's assertion that its objective is merely to prevent Wright from "lining his pockets" with profits from his book sounds like humbug; without the publicity from all the legal actions, it is unlikely that Wright would have earned an estima-

ted \$250,000 in royalties during the first ten days of his book's publication in the United States. In this case the British cabinet would be better employed seeking ways of preventing future breaches of confidentiality by members of its intelligence services. Its lawyers, who are said to have spent £1 million already in their pursuit of Wright, would be better employed drafting watertight contracts with his successors. If, as seems possible, it is a problem that people wishing to spill the beans move outside their governments' jurisdiction (there are some recent cases in the United States as well as in Britain), is it improper to think of some form of extradition treaty between like-minded governments? Better that than censorship. □

Going for broke

The largest British research council is in a hurry to set up novel research centres. Will they succeed?

THE British Advisory Board for the Research Councils (ABRC) has not had to wait long for a reaction to its suggestion that research councils should in future support academic research at British universities by setting up goal-oriented research centres near but not within universities. The scheme, first described only three weeks ago (see *Nature* **328**, 280; 1987), and endorsed only cautiously by the government, seems to have been embraced enthusiastically by the largest of the research councils, that responsible for science and engineering (SERC). Now SERC says that it hopes to set up no fewer than seven of these centres, not to mention that on the science and engineering of superconductivity for which bids have already been invited (see *Nature* **328**, 370; 1987, and page 464 in this issue). To the extent that these developments promise change in an otherwise stagnant setting, they are to be welcomed. But there is a case for asking that SERC will tread carefully in what it is about.

Even in Britain, university-based research centres are not unprecedented. The Medical Research Council (MRC) will, for example, quickly say that it has for more than 30 years been doing what SERC now plans, although a better analogy is the network of university research organizations founded over the past two decades by the Wolfson Foundation.

What SERC is now planning (with the collaboration of the Agricultural and Food Research Council in the case of a centre for process technology), are centres which are intrinsically interdisciplinary, but whose academic staff will include many from the host institution (polytechnics as well as universities are allowed) or institutions (consortium bids will be encouraged). There will also be research staff provided centrally, on longish but not permanent contracts, a director with "power" to change the research programme in the light of changing circumstances and a management committee to ensure that he or she has a programme. Industrial collaboration is talked of.

So far, so good. SERC's good intentions will not be questioned, but only the possibility that it may be trying to kill too many birds with the same stone. One objective is to shift the pattern of British research in directions that underpin industrial innovation, another is to encourage interdisciplinarity. That too makes sense. But the plan that the new centres should provide hospitality for visiting academics looks like an intended solution to quite a different problem, that of occupying the skills of able academics likely to be underoccupied except as teachers when the planned practice of selectivity begins to bite. SERC is plainly also ambivalent about the management of its centres. While warning academic institutions in its invitation for bids that some of the new ventures may fail, it plans to keep a hand in their management as if persuaded that it can help avoid that happening. Would it not be wiser to stand well aside? Seeing whether academic enterprises can sink or swim should be another part of what is a bold experiment — much bolder than that being mounted by the National Science Foundation in the United States. □

Expedition faces Antarctic ozone hole mysteries

- Aerosol fears smooth way for cooperation
- October deadline for results

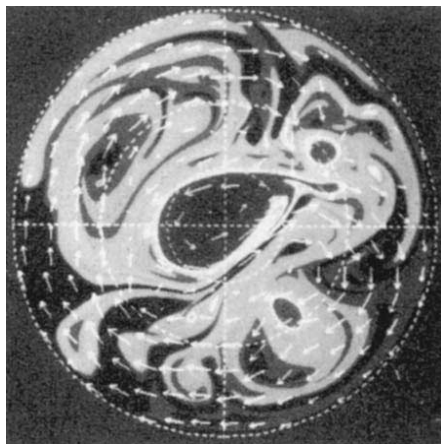
Washington

ANTARCTICA'S unexplained 'ozone hole' is to be investigated in a large-scale international expedition running from mid-August to the end of September. The aim is to find out whether natural processes or man-made chemicals are to blame for the sudden depletion of ozone in the stratosphere during the Antarctic spring.

The expedition will not see many explorers trudging across the lonely Antarctic ice. It will be based in the large city of Punta Arenas in southern Chile and most observations will be made from specially instrumented high-flying aircraft, including a modified U2 spy plane capable of reaching the centre of the ozone hole 18 km above the Earth.

Quick work was needed to put the expedition together. The ozone hole was discovered just two years ago by British Antarctic Survey scientists, but the US National Aeronautics and Space Administration (NASA) has succeeded in obtaining the cooperation of the governments of Argentina, Chile, France, New Zealand and Britain, and help from other research organizations such as the Chemical Manufacturers Association. The project's urgency stems from fears that the release of man-made chlorofluorocarbons (CFCs) lies behind the dramatic appearance of the ozone hole.

Just a few weeks ago, an international convention agreed in Vienna that CFC release must be cut back lest ozone levels fall and an increasing amount of damaging ultraviolet light reach the Earth's surface.



They also serve who stand by their computers — this picture, generated using a Cray, shows results of a high-resolution numerical model of the winter stratosphere developed by M.N. Jukes and M.E. McIntyre (to appear in next week's *Nature*).

Their worries stemmed partly from Nimbus 7 satellite data showing that global atmospheric ozone levels have fallen 2–8 per cent in the past seven years. But in the Antarctic, changes are far more disturbing: over a decade, spring ozone levels have fallen 50 per cent in the whole air column and 80 per cent at the centre of the ozone hole in the lower stratosphere.

The Antarctic atmosphere is so special that there seems no likelihood that such dramatic changes will spread, but if man-made chemicals are to blame for the ozone hole then calls for tight international regulations are bound to increase. As yet, a natural explanation cannot be completely ruled out. Antarctic ozone levels always decrease in September, most probably because the stratospheric ozone absorbs radiation when the Sun reappears at the end of winter. The heated air mass may then rise and be replaced by upwelling air from the troposphere where ozone levels are much lower. Any factors that enhance these processes — even particles injected into the stratosphere by volcanic eruptions might help to absorb radiation — could in theory turn the dip in ozone concentration into a precipitous drop.

A chemical explanation already receives support from Antarctic measurements showing unusual concentrations of chlorine and nitrogen oxides. But expedition scientists are reconciling themselves to the possibility that the cause of the ozone hole may be a complicated mix of chemical and dynamical factors. The chemistry alone will be difficult enough. While the favoured explanation is that low levels of nitrogen oxides permit chlorine oxides of industrial origin to catalyse the destruction of ozone, it seems likely that man-made bromine compounds are also involved.

The key data for the expedition will come from the aircraft. A DC-8 will make ten trips at 12 km, in the lower part of the ozone hole, and measure ozone and aerosol profiles, chlorine dioxide, bromine oxide radical, nitric acid, chlorine, nitrate, nitrogen oxide and hydrochloric acid concentrations by air sampling and remote-sensing techniques. Two flights are planned to go all the way to New Zealand and back. Complementary, shorter flights will be made by ER-2, the modified U2 spy plane, at 20 km in the centre of the ozone hole.

Although the data are unlikely to provoke a simple answer, Robert Watson, NASA programme director, has com-

Monsoon dating depends on Cray

Bangalore

THE project to achieve accurate forecasting of the arrival of the Indian monsoon season — which would be of great value to India's farmers — is falling behind schedule. And Indian scientists say the reason is the United States' refusal to supply the supercomputer they want.

The joint India–United States monsoon forecasting project is already five years old and the need for a supercomputer to deal with the enormous quantity of data that has accumulated was recognized in 1984. After a study of the programme's needs, Indian scientists at the Meteorological Department asked to purchase one of Cray's top models, the X-MP-24. The US side responded by offering the Cray X-MP-14, a slightly less sophisticated version, even though the original choice had been made after consultation with the US State Department.

Political analysts in New Delhi believe that the possibility that India might use the supercomputer for defence research, or even that the technology might reach the Soviet Union, had alerted hawks in the US State Department and the Pentagon. The State Department points out, however, that India is the first country to be offered a machine as advanced as the X-MP-14.

An Indian technical mission is now once more re-evaluating the project's needs to see if they can make do with the Cray X-MP-14. If not, they may have to consider the purchase of a Japanese or even a Soviet supercomputer, although opinions so far are that neither country produces a computer to match the Cray.

Radhakrishna Rao

mitted the expedition to producing a report of their findings by 1 October — a mere two days after ten weeks of aircraft observations ends. To make that possible, theoreticians are to be among the 160 scientists, pilots and technicians going to Punta Arenas and will try to build their models as the data arrive. Watson freely admits that, at worst, the report "might be a half-page long" but believes the project's significance to be too great to allow scientists a year to mull over their results.

Even with that deadline, the announcement will come just after the meeting at Montreal at which international guidelines on CFC use are likely to be decided. But if the Antarctic data are relevant, changes can always be made later.

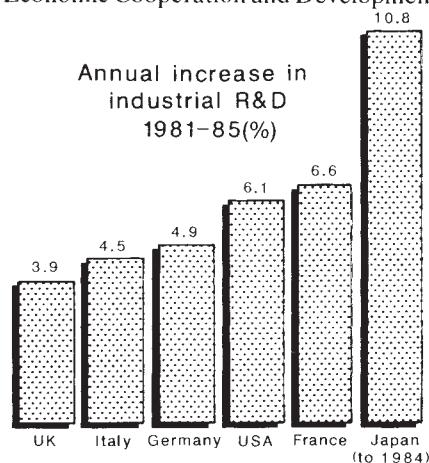
Chlorine compounds are virtually indestructible in the atmosphere and, as Watson points out, even if their production were halted tomorrow, it would be a century or two before they disappeared from the stratosphere. Alun Anderson

Short-term thinking behind dismal UK research spending

London

YET another report has been issued pointing out British industry's dismal performance in funding for research and development, this time by the Technology Requirements Board of the Department of Trade and Industry (DTI). But underlying the expected recommendations on increased support for "innovation" is the need for a fundamental change in the attitude of financial backers, away from a short-term profit-oriented outlook.

Industry-financed expenditure for research and development in the United Kingdom has grown at a lower rate than in other industrialized nations in the past 15 years (see figure), a fact reflected in a decline in patenting activity: the UK share of patents taken out in the United States has fallen by one-sixth since the mid-1970s. Figures from the Organization for Economic Cooperation and Development



(OECD) also show that Britain's trade balance in high-technology products has deteriorated sharply.

The root of the problem lies to a large extent in the City, Britain's equivalent of Wall Street, where an attitude of 'short termism' prevails. Company shareholders, largely insurance companies, unit trusts and pension funds, expect higher short-term profits than in other countries. Investment in research and development can soak up immediate profits, resulting in a loss of interest income. Britain's high interest rates (11.6 per cent, compared to 7.5 in the United States, 6.5 in Japan and 5.4 in West Germany, according to OECD figures for 1985) mean that it is difficult for investment in research and development to compete. With the emphasis on profits rather than the expansion that would result from innovation, debate centres around short-term versus long-term benefits, and value for money in research and development.

Scientists and industrialists argue that shareholders must put a premium on

growth and take pride in a company's expansion. They say that whatever the short-term profit level, a company that does not innovate must eventually die. "The short-term views of the City can't be used as an excuse for poor management", says Dr Penny Birdseye, deputy director of the Confederation of British Industry. "There is evidence that if the advantages of innovating are communicated well to shareholders, the support will come."

This week's report calls on industry to take advantage of Britain's high gross domestic product (GDP) growth rate by investing in long-term prospects for creating wealth. It recommends that industry should increase its support for research and development by 5 to 10 per cent a year, instead of the current 3.9 per cent.

Disclosure of research and development expenditure in company accounts is also recommended as a way to convey future prospects to investors. The government has supported this concept in a recent white paper, threatening to legislate if voluntary compliance is not forthcoming.

The technology board argues for a sharp increase in DTI support for innovation in industry, currently £387 million a year. It suggests that expenditure should focus on technology for information, advanced manufacturing, materials and biotechnology, and recommends encouragement of collaborative international projects such as the European Eureka programme and the LINK initiative for cooperation between industry and the universities.

Increased government support, not surprisingly, is another recommendation, especially for civil industrial development. In 1985, government support in this area, as a proportion of GDP, was 50-100 per cent higher in Germany, Italy and France than in the United Kingdom. British government funding for civil industrial development was 0.088 per cent of GDP in 1985, compared to 0.68 per cent for defence research and development.

One-quarter of military procurement spending goes for research and development, but the government is concerned that this may be too much. In its recent white paper, it announced the intention to reduce defence research and development spending because it may be taking scientific, engineering and skilled manpower resources from the civil sector. But this conclusion is not unanimously supported — some industrialists say such a move would harm Britain's balance of payments and threaten employment.

While industry pushes for more government palliatives to improve spending on

Bids requested for centres of excellence

London

THE first step has been taken towards a major reformation of research in British universities, with a letter sent out by the Science and Engineering Research Council (SERC) to all vice-chancellors and principals, inviting bids to set up interdisciplinary research centres.

The council intends to establish two or three centres annually over the next few years, to provide a focus for research in areas that cut across the normal university department structure and that are considered of strategic importance, including surface science, semiconductors and novel materials, molecular sciences, lasers in manufacturing, high-temperature superconductivity, engineering design, and process simulation, integration and control.

The new centres will be research groups in their own right, carrying out their own programmes of basic, strategic and applied research, advised by a management committee that will include academics and industrialists. Industry may also be involved in financing the centres.

SERC says it has no prejudgements about which higher education institutions will host the new research centres. Even the decisions about which topics will be focused on are not necessarily set in concrete. "A learning process is involved", says SERC chairman, Professor E.W.J. Mitchell. Financial support is also an uncertainty, and Mitchell acknowledges that the programme may face a short-term financial crisis.

Each inter-university centre will be based at one campus, possibly serving a group of universities and polytechnics in one locality. Each is expected to have a staff of about 40, comprising 15 technicians and 25 research and support staff.

Bids from interested institutions are due by the end of September, and SERC hopes to reach some decisions early in 1988. Highest priority is being given to a centre on high-temperature superconductivity, and an announcement on its location is expected before the end of the year (see *Nature*, 328, 370; 1987). Kathy Johnston

innovation, the government is calling for industry to stand on its own feet. Industry points out that it is difficult to make changes overnight, and asks the government to help by bringing interest rates down and by encouraging a good scientific base, thereby increasing the number of scientists available to work in research and in finance houses. Industrialists have set up a task force with representatives from the City, with a report expected later this year. There are signs that all the discussion and debate is beginning to bring a change.

Kathy Johnston

Grants for Japan's top twelve with physicists favoured

Tokyo

JAPAN'S Ministry of Education, Culture and Science (MESC) will give out more than ¥2,500 million (\$17 million) in "special, distinguished grants" to a handful of researchers this year, up by a massive ¥700 million from last year. As in previous years, molecular biology features prominently, taking five of the twelve awards. But the biggest grants go to high-energy and solid-state physics and the new field of high-temperature superconductors.

The awards are the most significant and lucrative in MESC's repertoire and are intended for "internationally recognized research likely to produce outstanding results". Each award runs for 3–5 years and is largely used for the purchase of equipment and materials. The salaries of the half dozen or so key researchers on each project are paid separately by the ministry.

Peptide research is a favourite this year. Shigetada Nakanishi of Kyoto University, who receives ¥204 million over four years, will investigate regulation of bradykinin, neuropeptides and the angiotensin-renin system both *in vitro* and *in vivo* (using transgenic mice) and will try to clone the peptide receptors. Hisayuki Matsuo of Miyazaki Medical College gets ¥223 million, also for four years, to search for peptides involved in "bio-communication" (the chemical transfer of information within organisms). Similarly, cell signalling will feature in a three-year study

of GTP-binding protein by Yoshito Kaziro of Tokyo University.

But the biggest grant in bioscience goes to Nobutaka Hirokawa of Tokyo University to study the cytoskeleton, in particular of nerve cells. Hirokawa, who gave up a university post in the United States to return to Japan, is one of the pioneers of quick-freeze electron microscopy and he will use this technique and micro-injection of proteins and antibodies to examine the molecular architecture and dynamics of the cytoskeleton, and localized developmental processes in the cell.

Physicists, however, run off with the biggest grants of all. Yoichi Iwasaki of Tsukuba University gets ¥278 million over three years to calculate the masses of hadrons, such as the H particle, using numerical simulation based on lattice gauge theory. And Noriaki Ito of Nagoya University receives a similar amount over five years to probe alkali halides, semiconductors and ceramics with lasers and electron beams in an attempt to elucidate electron-photon interactions.

But in solid-state physics, the award of ¥258 million to Shoji Tanaka of Tokyo University is attracting the most attention. Tanaka heads the group at Tokyo University that confirmed high-temperature superconductivity in copper oxide ceramics at the end of last year, following their discovery by scientists at IBM Zurich. And the Ministry of Education is making considerable financial commitments to this group's research. Earlier this year, the ministry took the extraordinary step of extending a special project grant headed by Professor Nakajima, one of Tanaka's colleagues, by injecting a further ¥160 million (over \$1 million) for the study of oxide superconductors. And Professor Kazuo Fueki, another of the Tokyo University group, was also awarded ¥36 million for this fiscal year.

Further substantial funding from the ministry (for high-temperature superconductor research) can be expected. Simultaneously with the announcement of this year's special distinguished grants, the ministry released details of projects selected as "priority areas of research" in fiscal year 1988. Among them is one on high-temperature superconductors headed by Yoshio Muto of Tohoku University, who also has close links with the Tokyo University group.

Although the number of participants and amount of funds have yet to be fixed, about 60–70 researchers are expected to join Muto's project and funding will probably be in the middle of the range for such projects (¥50 to ¥600 million), according to a ministry official. David Swinbanks

Superphénix on hold

THE French prototype Superphénix fast-breeder reactor is still in trouble. The leak of liquid sodium from a storage tank that was first noticed in March has yet to be located. The tank will now be drained and a search made for the leak by pumping helium into the space between the tank and its outer protective shell. Auditory detectors should pick up the escaping gas.

Officials at the Superphénix site, near Lyons, strongly deny that faulty workmanship is responsible for the leak. Mr Horst Weber, a construction foreman at Superphénix ten years ago, had claimed in the German magazine *Stern* that welding work had sometimes been botched. P.C.

More animals spared

THE number of laboratory experiments using animals is declining in Britain, according to the latest statistics, and is now the lowest for 30 years. About 80 per cent of the 3.1 million animal experiments carried out in 1986 involved mice or rats, and the number of primates used has dropped by about a third in the past decade, to 5,600 experiments, mainly for studying medicines. New legislation should further reduce animal experimentation. K.J.

Acid rain deposition

ACID rain deposition in the United Kingdom declined in the first half of the 1980s, according to a new report from the UK Review Group on Acid Rain. But levels of hydrogen and sulphate deposition in the high-rainfall areas of Cumbria and the Scottish Highlands remain high, comparable with those in the acidified areas of Norway. Emissions of sulphur dioxide have declined overall, with a proportional reduction in deposition. Hydrogen-ion deposition has dropped by 25 per cent, although nitrogen oxides have not changed since 1970.

The review group, made up of representatives from universities, industry and the government, says good quality data were not generally available, although two new monitoring networks have recently been established. K.J.

AIDS tests in India

THE Indian health ministry has announced that all foreigners intending to stay for more than a year in India will have to undergo tests for AIDS (acquired immune deficiency syndrome). The rule does not, however, apply to diplomats and their staff or to journalists.

AIDS screening is at present compulsory only for foreign students. Official sources said that the decision to extend this to all foreigners staying longer than a year was taken to placate the foreign students, mostly Africans, who have been complaining that India's AIDS policy was discriminatory. K.S.J.

New angle on flight



THIS is not a mistake, but what the US aviation company Rockwell intends to do to an F-8 Crusader aircraft. An oblique-wing aircraft, according to aerodynamics theory, might not create a sonic boom at low supersonic speeds. Drag is reduced too, as the lift of the wing is distributed over the long streamwise length. Finally, a boon in naval applications, an oblique-wing aircraft would be much more convenient for stowage in below-deck hangars.

National Science Foundation taking a bipolar perspective

Washington

THE National Science Foundation (NSF) is formulating plans for a new initiative for polar studies. Called Arctic System Science (ARCSS), the new programme, planned to start in 1989, will take a multidisciplinary approach to Arctic studies, in much the same way as the International Geosphere/Biosphere Program will examine a variety of Earth systems in studying global change.

The NSF has given a lot of attention to the Arctic in recent months. The Arctic Research and Policy Act of 1984 created the Arctic Research Commission and the Interagency Research Policy Committee, both chaired by NSF. The committee has now produced the 5-year US Arctic Research Plan, signed by President Reagan and soon to be transmitted to Congress.

A special task committee of the National Science Board has also just completed a report of NSF's role in the polar regions. Chaired by Board member Rita Colwell, the report recommends that science — rather than available logistics — should dictate Arctic programmes. Because the poles are difficult environments to work in, it is important that research plans accommodate as many intersecting inter-

ests as possible. The Colwell report urges NSF to taking a leading role in the development of polar activities among interested US agencies.

The Arctic is looked on as a sensitive indicator of changes in the global climate — in fact it may not only reflect global environmental changes, it may drive them. Melting ice from the poles may release methane, another greenhouse gas, and that could increase global warming in a positive feedback system.

The ARCSS initiative is intended to start a number of projects that will study the forces controlling the polar environment. One project, Arctic land-air interactions, will attempt to study how changes in Arctic ecosystems are driven by the global climate. Another, high-latitude oceanic and atmospheric processes, will examine how energy exchanges alter circulation patterns in the Arctic Ocean.

ARCSS is currently under internal review at NSF as the agency prepares for negotiations with the White House Office of Management and Budget over its 1989 budget. But with the often repeated intention of doubling the NSF budget over the next five years, new initiatives like ARCSS appear to have a good chance of being adopted. **Joseph Palca**

Aurora borealis pair for Soviet launcher?

Garching, West Germany

A GROUP of Western European space scientists are turning to the Soviet Union for a rocket to launch their next satellite experiment. The project, called "Turbulence", involves putting two satellites into polar orbits in 1993 in an international project to study the physics of the aurora borealis.

Sweden, West Germany and Great Britain are expected to share the estimated DM100 million cost of the two satellites. The Soviets and others will be invited to contribute telemetry equipment and make measurements of their own using the satellites. Gerhard Haerendel of the Max-Planck Institute for Extraterrestrial Physics at Garching, who is one of the initiators of the project, says that using a Soviet rocket would cut costs. An Ariane 4 rocket, built by the European Space Agency, would cost 80 million ECU (European Currency Units, 1 ECU=0.70), said Haerendel, and there would be "no chance" of obtaining this money from the governments in question. The Soviet Union will provide the rocket for nothing, Haerendel said, in exchange for "data like everyone else gets" as well as the prestige of launching a Western satellite.

Soviet approval of the project is un-

official as yet, but Academician Roald Sagdeev, science adviser to Mikhail Gorbachev, has "promised" Haerendel that it will go on as planned. Sagdeev, himself a plasma physicist, is director of IKI, the most influential space science institute in the Soviet Union.

Only two satellites have been launched so far to study primary aurora borealis phenomena — the S3-3 of the US Air Force and the Swedish Viking satellite. Viking, launched in 1986, has already made impressive observations of the exchange of electric current between the magnetosphere and ionosphere, currents which arise in part due to the solar wind. Japan is due to launch a satellite called EXOS-D in 1989 which will also be able to inject charged particles into the magnetosphere and detect them after they "boomerang" back to the spacecraft. Turbulence hopes to go further and fire electrons and charged particles between two spacecraft oriented along the same magnetic field line.

Researchers in the three countries involved are in the preliminary stages of applying for funding for Turbulence. Haerendel said he expects approval "not before mid-1989". **Steven Dickman**

Soviet trio in US Congress

Washington

IN an unusual move, three Soviet scientists travelled to Washington last week to testify before the new US House of Representatives subcommittee on international scientific cooperation. A biologist, a glaciologist and an atmospheric scientist from the Soviet Academy of Sciences brought a message to Congress stressing the urgent need for international participation in the International Geosphere/Biosphere Program (IGBP) to study global change.

IGBP was launched at a meeting of the International Council of Scientific Unions (ICSU) last year, as a way to commemorate the 25th anniversary of the International Geophysical Year programme, and to expand upon its successes. ICSU is coordinating the programme, and the United States, the Soviet Union, Sweden, Canada, France, Japan and the People's Republic of China have agreed to join.

In July, the ICSU special committee for IGBP met in Paris, and established three working groups dedicated to global modelling, data handling and exchange, and the reconstruction of past changes in the Earth's environments. Harold Mooney, chairman of the committee representing the United States to ICSU, says planning is expected to extend until 1991, with the programme scheduled to begin in 1992.

The US efforts in IGBP will be divided between several agencies — research programmes are under way or are being planned that will tie into programmes by the National Science Foundation (see this page), the National Aeronautics and Space Administration and the National Oceanic and Atmospheric Administration, among others.

Two of the Soviet scientists who testified at the hearing, V.M. Kotlyakov and V.A. Troitskaya, are part of the ICSU committee for IGBP. Their testimony focused on past successful smaller international science collaborations, such as the Unesco Man and the Biosphere Programme, and the possible technical and political stumbling blocks that could stand in the way of IGBP.

The Soviets spoke in English, with only occasional help from an interpreter, and language and cultural barriers were still evident. But the Soviets' message, as given by the third scientist, V.E. Sovolov, carried through. He said "survival and development of each and all is a task that can be fulfilled only by all humanity, no people and no state being able to tackle it single-handed". The Soviet Union will host a meeting on IGBP in August, 1988, to discuss data exchange. **Carol Ezzell**

UK government urged to drop 'annuality' requirement

London

THE British government is being strongly urged to relax spending controls over the research councils to enable them to operate within a more flexible financial regime that would allow more efficient management of resources and encourage new kinds of commercial involvement. The government has been presented with two unpublished reports that highlight the problems posed when the research councils are faced with traditional Treasury restraints, particularly the system of 'annuality', which requires the councils to spend all their income from government and non-government sources within the financial year in which it is received.

The Advisory Board for the Research Councils (ABRC), whose principal job is to recommend the distribution of funds among the five research councils, commissioned a study from Sir Ralph Riley, a former secretary of the Agricultural and Food Research Council, to examine the ways in which Treasury controls were frustrating research councils' efforts to secure value for money from their funds. Similar issues have been addressed by a resource management study, undertaken by the Department of Education and Science and the research councils.

ABRC says that if the government wants public institutions to be more business-like, "the traditional public accounting limits and controls, on the councils in particular, need to be correspondingly redesigned and reduced".

The annuality problem becomes particularly acute with earnings from non-

government sources, which are usually paid in arrears, where liquidity is needed to hire staff, initiate work and carry it through to the first payment. Riley recommends that councils be allowed to build up reserves from their commercial earnings.

Annuality has always been a thorn in the side of many public bodies, but ABRC says that if the research councils are to respond to stated government policy of



Sir Ralph Riley, author of the ABRC report.

increasing earnings from external sources, some relaxation of this rule is essential. The Treasury will almost certainly attempt to resist such moves, despite its contention that it is not "in principle" against the annuality rule. The annuality rule, says the Treasury, was arrived at by "bitter experience", and is the only realistic way to provide a framework in which to operate spending control and allow accurate forecasting, given that public expenditure as a whole is determined on a yearly basis. It points out the flexibility that already exists, including a 3 per cent annual leeway for the Medical Research Council's grants provision and 4 per cent for international subscriptions, to allow for fluctuations in exchange rates. "But you can go too far down that road", said an official, "there is only so much room for latitude."

Simon Hadlington

Final blow to UK space plan

London

HOPES for an expanded British space effort have been dealt a severe blow with the announcement that the government will not provide any new money for the British National Space Centre (BNSC), a body set up in 1985 to coordinate the nation's space policy. The decision will result in Britain having less of a stake in the European Space Agency (ESA), and will almost certainly mean a reduction in the domestic science programme. BNSC's last hope lies with the newly formed Advisory Council on Science and Technology (ACOST), which meets in September.

Last year, BNSC presented to the government its 15-year national space plan, which remains unpublished but is believed to have requested an annual budget of £300 million, bringing Britain's spending on civilian space to about half that of France or West Germany. BNSC itself does not directly receive funding from the Treasury; instead, its budget (£112 million this year) is made up of the space funds of the Department of Trade and Industry (£64 million), the Science and Engineering Research Council (£30 million), the Ministry of Defence (£17 million) and the Natural Environment Research Council (£1 million). About £80 million goes to ESA, with the remainder on the domestic programme.

BNSC had hoped that the extra £200 million would come from central government. Now it is faced with the almost impossible task of trying to win a larger share of the existing civil research and development budget of £4,500 million.

In response to a question in the House of Commons on 23 July, Mrs Thatcher said that no new money would be forthcoming "at present", and that the relevant ministries had not agreed to transfer resources from existing programmes. She then suggested the by now familiar panacea: "I hope that the private sector, if interested in the results of such research, will come forward with considerable resources."

Industrialists and academics have been angered by the decision. BNSC officials are putting on a brave face and saying that the space plan has not yet been rejected outright. Space experts say that Britain needs at least £200 million a year to maintain a viable domestic programme and benefit from international collaboration.

Earlier this year, BNSC's director general, Roy Gibson, concluded a lecture by asking the government an open question: "We have the competence, we have the experience and we have the necessary enthusiasm and faith — may we please have the money?" Mrs Thatcher has answered.

Simon Hadlington

Restructuring worries Soviet scientists

London

MR Mikhail Gorbachev's policy of *perestroika* (restructuring) — in science as elsewhere — assumes that the provision of improved material incentives, in particular bonuses and proficiency supplements, will ensure the high-level work the country needs. But a recent survey carried out by the Byelorussian Academy of Sciences throws some doubt on this assumption. Of the scientists polled, 48.5 per cent considered that pay had no major effect on their research performance, and 13 per cent thought that it had no effect at all.

"Interest in science" came only fourth in the scientists' ranking of motives for scientific activity. Eighty-four per cent put "conscientious work habits" first, followed by "responsibility to the collective" and "the desire to work for the good of society". Dr Evgeni Babosov, director of the academy's Institute of Philosophy and Law, said on

Soviet television that without a "cognitive and deep interest", the "creative impulse in science" cannot exist. Moreover, the lack of interest in bonus payments suggests that the restructuring of pay scales is simply not working properly.

In many institutes, adjustment of the pay structure has simply meant giving more of the harder-working scientists a flat rise of 20 or so rubles a month and penalizing the less industrious by a comparable amount. Babosov did not indicate whether the poll was carried out anonymously. Gorbachev's reforms are directed against slackness, shirking and egoism "Conscientiousness" and "serving society" might well be felt to be appropriate responses — particularly among those attracted into science by its social status, who now feel themselves in danger of being "restructured" into some less valued profession.

Vera Rich

US Navy oceanography satellite heading for a comeback

Washington

GIVEN up for dead late last year, the Navy Remote Ocean Sensing Satellite (NROSS) appears poised for a comeback. Navy officials have told Congress that they would like to restart the programme, and are now trying to gather together funds that had been redistributed by the Pentagon when the project was axed.

The Navy had originally proposed NROSS because of its importance for sea-surface observations. Most operational meteorological satellites carry infrared or visible radiometers that give information about clouds. To pierce the clouds, the Navy developed a low-frequency microwave radiometer, capable of providing all-weather measurements of sea-surface temperature. At high northern and southern latitudes, this ability is critical, because the ocean surface is hidden by clouds 70 per cent of the time.

Besides the radiometer, NROSS would have carried an altimeter for measuring wave heights and an instrument called SSM/I (special sensor microwave imager) that provides information about ice concentration and water vapour content. A sophisticated scatterometer capable of providing information about both wind speed and direction would also have been on board.

But NROSS faced budget problems, and in December, the Navy decided to cancel the project. The cancellation provoked a flood of protest. James Fletcher, director of the National Aeronautics and Space Administration (NASA), and Erich Bloch, head of the National Science Foundation, both wrote to then Navy Secretary John Lehman urging the retention of the programme. NASA's interest in NROSS was pragmatic, since NROSS' scatterometer is a NASA instrument being built at the Jet Propulsion Laboratory. Data from the scatterometer, called NSCAT, will be shared between the Navy's operational needs and NASA's research aims.

Lehman's initial decision to cancel the project was based on an estimated total cost of \$420 million, and NROSS supporters in the Navy argued that the true cost was closer to \$350 million. On 10 April, Lehman wrote to Congress announcing his intention that the project be restarted. The main Department of Defense must also be convinced if NROSS is to survive. A formal review by the office of the Secretary of Defense is just getting started, and should be completed this fall. If the Pentagon gives its blessing, NROSS could have a late 1991 or early 1992 launch.

Joseph Palca

Putting more "I" into Australia's CSIRO

Sydney

FURTHER change is on the way for the Commonwealth Scientific and Industrial Research Organization (CSIRO), Australia's largest research body. Eight months ago a shake-up at the top gave CSIRO a new corporate-style board (*Nature* 324, 608, 1987). Now the whole organization is to be restructured to make it more responsive to the needs of industry.

CSIRO employs 2,500 scientists and has an annual budget of A\$450 million. Under the new scheme its 41 divisions grouped into five subject areas are to be cut to 32 divisions in six areas. The reorganization plan comes from a management consultancy, McKinsey and Co., the choice of which as an advisory body says much about the new board's thinking. CSIRO will no longer be run by scientists for science's sake but will take on more of a business philosophy, and less reliant on public money.

Attempts will also be made to cut red tape by decentralizing administration away from CSIRO headquarters in Canberra and placing more authority in the hands of the divisional and institute chiefs. They will receive a crash course in man-

agement techniques while 120 jobs will be lost at CSIRO headquarters.

The move towards industry is partly dictated by necessity. CSIRO's budget had gradually been eroded while research costs have risen, according to Dr Keith Boardman, the organization's chief executive officer. While CSIRO once conducted up to 2,000 research projects the number is now nearer 500. He hopes the present A\$55 million received from industry will grow to A\$100 million by 1990.

McKinsey and Co.'s new structure divides CSIRO into "business systems" which integrate research along the same lines as the industry it will serve. There is, for example, a natural sequence for wool-related research which will be grouped administratively to include pastures and soil research, sheep reproduction and nutrition, wool harvesting, wool processing and textile manufacture. Divisions with little industrial relevance fear even leaner times than at present. But some reassurance has been gained from CSIRO chairman Mr Neville Wran QC, who has said that CSIRO remains committed to excellence in long-term and fundamental research.

Charles Morgan

Japan boosts rocket power

Tokyo

WITH a request to the Space Activities Commission submitted on Monday, Japan's Institute of Space and Astronautical Science (ISAS) began manoeuvring for permission to build a rocket that would one day be capable of carrying scientific missions to the planets. The request, for a study of a solid-fuel rocket to triple ISAS's payload capacity, would break a political argement that has limited the size of ISAS's rockets for the past 17 years.

The agreement came into being with the creation of the National Space Development Agency (NASDA) in 1970. NASDA was given responsibility for applications satellites and bought advanced US technology while ISAS remained an academic organization launching scientific probes and satellites with its own home-developed solid-fuel rockets. A price for the continuing independence of ISAS (then a part of Tokyo University) was an agreement that its rockets should not get much bigger — a limit of 1.41 metres tail diameter was set.

That limit still stand but ISAS engineers have shown extraordinary ingenuity in working within it — and all with a budget less than one tenth that of NASDA. Payload capacity has been built up by making the rockets taller and adding strap-on boosters. The latest model, the Mu 3S-II, cheekily bulges beyond 1.41 at the cone.

Since Japan's first satellite *Ohsumi* was placed in orbit in 1970, ISAS has successfully launched 18 more satellites and probes and has only had one failure (an X-ray satellite in 1978). And in 1989 ISAS plans to launch a lunar probe, Muses-A, that will swing around the Moon, perhaps several times.

But with Muses-A, ISAS's engineers are reaching the limits of their ingenuity. Wherein comes the proposed new rocket. The new launch vehicle would be capable of placing a 2-ton satellite in low Earth orbit — treble the payload capacity of Mu 3S-II, and slightly less than NASDA's H-I. But, although only a couple of metres taller than Mu3S-II, the 1.41-m tail limit would have to be breached.

The Space Activities Commission, which formulates Japan's space policy, will consider the ISAS request over the coming months. ISAS has already drawn up a design for the new rocket and, if given the go-ahead in time for the fiscal 1989 budget applications (which are made a year from now), it could be ready for launch in 1993. Among the missions under consideration are one to Venus, which could involve the release of a balloon into the Venusian atmosphere, and a follow-up to Muses-A.

David Swinbanks

Glaxo's new acquisition signals Biogen retreat from Europe

London

ONCE-PROUD multinational Biogen, the biotechnology company with bases at Geneva and Cambridge, Massachusetts, but with its registered office in the Netherlands Antilles, almost withdrew from Europe last week with the sale of its Geneva laboratory to Glaxo, the British-based pharmaceutical company. Biogen retains an interest in Bioferon, a joint venture in West Germany, and in a protein engineering laboratory in Belgium, but is otherwise concentrated in the United States.

This development has been driven by financial considerations, especially by Biogen's high ratio of operating costs to income during the past three years. The Geneva laboratory was the first of the company's research facilities. It has been sold to Glaxo for an undisclosed sum, but as a going concern. Glaxo will take over the existing staff of about 100, together

with responsibility for two products now in clinical trials (interleukin-2 and GPD-6) and worldwide market rights except in Japan and some neighbouring countries.

Glaxo says openly that its objective is to make faster progress in the development of new pharmaceutical products than would otherwise be possible. A spokesman estimated last week that the Geneva laboratory would put the company three years further ahead than would have been possible by internal growth.

The company plans to keep the Geneva laboratory as a separate facility, but the research programme will be set by Dr Allan Williamson, who left his post as a Glasgow professor to become director of scientific research at Glaxo.

Among drug companies, Glaxo has recently been among the most profitable. In the financial year to the end of June 1986, its pre-tax profits were £611 million on sales of £1,400 million. □

French lose medical imaging capability

Paris

IN a deal that has caused concern among French medical researchers, the nationalized electronics company, Thomson, last week exchanged its medical imaging subsidiary, CGR (Compagnie Générale de Radiologie), for the domestic electronics division of the United States' General Electric. Some opponents of the agreement feel that the loss of CGR will bring to an end the previous close involvement of French medical researchers and practitioners in the design and manufacture of specialist radiological equipment.

CGR and its subsidiaries specialize in the therapeutic use of radiation and supply 60 per cent of radiological equipment found in French hospitals. For several years, research teams from INSERM (Institut National de la Santé et de la Recherche Médicale), France's principal medical research body, have contributed marketable ideas to CGR, while also being major clients. In 1985, INSERM and the Centre National de la Recherche Scientifique (CNRS) ordered a multi-million-dollar cyclotron from CGR for their joint medical laboratory at Lyons, without calling for international tenders, so sure were they of its appropriateness.

CGR's involvement in radiology will continue under General Electric's control, but developments in magnetic resonance imaging and scanner technology will be transferred to the parent company, depriving France of its place in these newer markets.

Peter Coles

Europe agrees on car exhaust emissions

London

AN agreement reached last month on car exhausts in Europe looks set to become a test case to clear the air surrounding the European Economic Community's (EEC) decision-making process.

The directive setting limits for vehicle emissions had been stalled for two years because of opposition by Denmark, which sought stricter standards like those in the United States. But under new EEC voting provisions designed to speed up decisions, the package was approved by eleven member states, overruling Denmark.

Denmark is now expected to adopt tougher standards within its own borders, allowed under the new voting rules on environmental protection grounds. But this would create a barrier to the free trade of vehicles, and the whole issue is likely to be taken by the EEC to the European Court of Justice as a test case.

The directive adopted last week divides vehicles into three categories, based on engine size; targets for large, medium and small cars respectively are 45, 30 and 25 grammes per test for carbon monoxide (compared to 16 in the United States); 15, 8 and 6.5 for hydrocarbons and nitrogen oxides (4.6 in the US). The relatively lax standards for smaller cars leave the door open for lean-burn engine technology.

The European Parliament is expected to approve the vehicle emissions package within the next few months, but prolonged legal wrangling could block the road to implementation.

Kathy Johnston

Reagan espouses superconductors

Washington

LIKE a commander-in-chief giving a pep talk to his troops as they prepare for battle, President Ronald Reagan, flanked by his Secretaries of Defense, State and Energy as well as his science advisor and the director of the National Science Foundation, told participants in the special Federal Conference on the Commercial Applications of Superconductivity that they were "among the pioneers carrying on [the] American dream into the future." The president also used the occasion to announce plans for a legislative package aimed at "preserving the competitive advantage of US industries" in the superconductivity field.

By bringing along his top cabinet officers, President Reagan appeared to underscore the importance his administration attaches to the commercial potential of the new superconducting materials. (An alternative explanation might be that he intended to draw attention away from the testimony of Attorney General Edwin Meese before the Congressional committee investigating the Iran-Contra affair.) Although the discoveries about superconductors have been coming from laboratories all over the world, the conference definitely had an "America first" flavour. Foreign participants were excluded from attending, although the foreign press was allowed to cover the event.

After a brief scientific session where Angelica Stacy of the University of California at Berkeley, Paul Chu of the University of Houston and Robert Schrieffer of the University of California at Santa Barbara discussed the brief history of the new generation of high-temperature superconductors, the conference turned to how these new materials might be used for industry.

President Reagan told the conference that he will propose legislation modifying antitrust laws so that US industry could cooperate in developing applications for the superconductors. The president would also like to see greater protection for proprietary information now available to all under the Freedom of Information Act. The administration also intends to modify the patent process to help improve competitiveness in superconductivity.

These legislative changes are part of an 11-point superconductivity initiative that will include a "Wise Men's" advisory group on federal policies relating to research and commercialization of superconductors. Other points include federal grants for new material processing schemes and \$150 million to the Department of Defense for superconductor research over the next three years.

Joseph Palca

Not cashing in on inventions

SIR—Your snippet of news under the heading "A helping hand for inventors", (*Nature* 327, 651; 1987) states that the "BTG scored a major triumph in the defence of British patents and realized substantial royalties for the British inventors of advanced medical scanners. The US electrical conglomerate General Electric is to pay BTG an undisclosed sum, believed to be millions of dollars, in royalties occurring from the sale of scanners in the past two years. Among the beneficiaries will be the teams from Nottingham and Aberdeen Universities."

As one of the members of the Aberdeen team, I can say that the impression created that the inventors will receive large sums of money is very misleading.

Sincere congratulations to the British Technology Group for their achievement in bringing the negotiations with the US General Electric Co. to a satisfactory conclusion: it is excellent news for Britain, the three universities and their departments involved, and it is good news for the scientists — but no more than good news and "millions" will certainly not come their way. On a machine sold at £750,000, the royalty paid by the company to BTG might be as much as £1,000. Allowing for overheads, including expensive patenting and high legal costs, the BTG take, the three universities' take and the university departments' take, it means about £5 to each of a team of six inventors from which about half will go in income tax. If 400 scanners are sold per year, this means £1,000 for one of the inventors to spend per year. This is better than nothing and very acceptable, but the inventor will have to be the proverbial cat with at least nine lives to make a million.

Also, does this not make clear another contributory factor to the brain drain? It was a brilliant piece of hard work; it is leading to an important and genuine advance in medicine; it was a 10-year fight to wrest the secrets from nature, to wrest the grants to make it possible against all the competition and the cuts. The BTG says in its press release that "the agreement will provide the originators of this pioneering work with financial returns commensurate with the importance of their contribution". Is the reward commensurate? One of the reasons leading to it not being so is that the patent laws in the United Kingdom are very favourable to the employer, while in the United States and West Germany they are more favourable to the inventor and less 'employer'- and institution-dominated. Could this not be a contributory factor to two of my team being now in the United States, and one in West Germany? The inventor has very few rights and hardly any say in what happens to his invention. It can be sold

and disposed of as company or institutional intellectual property without reference to him. Unless he can afford the very expert and expensive legal advice needed, he has no say or redress in the matter. However, we are left with the very important personal satisfaction of knowing that our salaries are coming back to Britain from abroad 20 times over per year. I hope everyone remembers that, too, in these days of university cuts.

JOHN MALLARD

*Department of Bio-Medical Physics
and Bio-Engineering,
University of Aberdeen
and Grampian Health Board,
Foresterhill, Aberdeen AB9 2ZD, UK*

Aids: another threat

SIR—The public health threat from the AIDS (acquired immune deficiency syndrome) virus compels attention to laboratory practice. It appears¹ that blood and possibly other body fluids from people with AIDS can infect others, either by intact mucosa or dermal cuts. These findings demand a review of many old, often comfortable, laboratory procedures. In my laboratory, we study the feeding behaviour of medicinal leeches; their potential for acting as an AIDS vector is viewed with dismay. In feeding leeches, we routinely use human blood obtained from a hospital blood bank where it is tested for both AIDS and hepatitis virus, but all feeding experiments must now be conducted while wearing rubber gloves, as AIDS is not always detected by the antibody test.

These developments face educational institutions with new responsibilities for their students. Students of introductory physiology courses, of whom there may be 500,000 a year in US colleges, are regularly exposed to the hazards from the blood and body fluids of others. These courses usually have laboratories which require blood samples be taken with lancets. The samples are used for cell counts, blood typing and differential staining. Blood is handled repeatedly as it is poured, smeared, dabbed, washed and centrifuged. Other experiments require handling and examination of fresh saliva and urine.

The potential for viral transmission among the large numbers of students as they conduct experiments in the close confines of the teaching laboratory is frightening. The risk of accidentally exposing the oral or optic mucosa to blood during sampling and handling is genuine. The hazards to instructors who clean up discarded lancets and slides is also not inconsiderable.

We must decide whether the risks

imposed by a potential exposure to deadly retrovirus are warranted by the "educational experience" of general teaching laboratories. At my university, we think not. All experiments in introductory physiology and biology courses that require students to be exposed to fresh blood, urine or saliva have now been cancelled (permanently, I suspect).

CHARLES M. LENT

*Department of Biology,
Utah State University,
Logan, Utah 84322-5305, USA*

1. III International Conference on AIDS, Washington, DC 1-5 June 1987.

Race relations

SIR—Since 1948, the number of deaths worldwide from racial strife is over a thousand times the South African total; and about 1,500 million people have sham votes in one-party states against 26 million without votes in South Africa. But condemnation of South Africa is much harsher than that of other countries. Clearly, racial discord and lack of suffrage alone cannot explain the ostracism.

Many white boycotters cannot admit that they have negative feelings towards blacks: three signs of this are excessive reaction, unconscious aggression and overcompensation. Any outside stimulus that stirs up such feelings is repressed as strongly as they are, hence the quite disproportionate reaction of boycotts and sanctions. The fact that these measures hurt blacks badly shows how repressed feelings continue to act — unconsciously. The overcompensation needed to keep the feelings out of consciousness appears in such ways as 'affirmative action' (which is also aggression against blacks because it insults them).

For the sake of good race relations may at least some boycotters gain insight soon.

J.A.D. EWART

*Delgany,
Solebridge Lane,
Chorleywood,
Rickmansworth, Herts WD3 5SW*

Immortality

SIR—I am sorry to see John Maddox once again rubbishing the tenets of Christian belief (*Nature* 326, 451; 1987, "immortality is impossible").

It is not that Christians are touchy about criticism, but he might mislead the ill-informed. The fact is, he does not have the data necessary to write off the 2,000-year experience of the Christian community.

It may be convenient for experimental purposes to 'exclude the principle' of this, but in so doing one can no longer make scientific statements about it.

JOHN DUDDERIDGE

*49 Macaulay Road,
Luton LU4 0LN, UK*

What is an electronic journal?

A newly published journal offers a prospect of what may be in store when people are persuaded of the folly of making hard copies so that journals can turn them back again into binary signals.

How is this for an extract from a journal's 'guide to authors'? "We can also process papers formatted with the `{\sc troff}/` system (which is usually provided with `{\sc unix}/`). Our procedure involves source-level machine translation from `{\sc unix}/` to `LATEX`. To ease this translation, avoid defining your own `{\sc troff}` macros. If possible, also avoid `{\sc eqn}` definitions. Try to use the `{\tt -ms}/` or `{\tt -me}/` macro packages."

Fair play, that is not so much an extract from the 'guide to authors' in the first issue of the journal *Complex Systems*, but the translation of that guide into the formatting code known as `LATEX`. The editor has had the happy wheeze of providing his guide in the form of a sample paper which he then helpfully turns into code so that authors "who intend to transmit your paper to us electronically" will know how to go about it. Then there is a list of no fewer than nine mailbox numbers, on systems such as Arpanet and Bitnet, with which authors may communicate.

The editor of this remarkable journal, Dr Stephen Wolfram, has been well known as an iconoclast since he walked out of an Oxford undergraduate physics course believing its content to be trivial, signing on instead as a graduate student at the California Institute of Technology. For a time, he was Caltech's favourite son, but, after a dispute about the rights and wrongs of selling a computer program privately, moved a few years ago to the Institute of Advanced Study at Princeton, where he has become the apostle of von Neumann's cellular automata (see *Nature* **311**, 419;1984), played a part in the development of the Connection Machine at Cambridge, Massachusetts, and has now, it appears, formed his own single-journal publishing company (Complex Systems Publications Inc.) as well.

The content of the first issue of *Complex Systems* fully reflects this background: almost all the papers are essays in the craft of cellular automata. Apart from the predictable demonstrations that cellular automata are marvellous ways of generating complicated patterns, there are some neat suggestions such as that of Puhua Guan from the University of Puerto Rico that cellular automata should be splendid ways of generating public key cryptograms (where the rules of encryption can be made public without much risk that intruders will be able to decode

messages) and for representing neurons in the sensory cortex by cellular automata in a way that allows for lateral connections between neurons and feedback between back and front. But the journal will be more interesting when the cellular automata are mixed with a more general discussion of complex problems.

That, however, is almost the least interesting remark to make about the new venture. One obvious and unanswered question is why an editor as committed as Wolfram to the computer as a way of life should have followed two very traditional practices in the design of his new journal — his willingness to accept contributions as typewritten paper, and his publication of the journal as inked paper bound between stout covers (rich red in this case). If that is what Wolfram believes is wise, what chance is there that other journals will go further, let alone as far? Time will tell, but the two traditional features of *Complex Systems* are well chosen. There is an important sense in which the proper use of the process of publication turns on two crucial questions of access. First, and most important, an entity cannot be said to have been published unless all who may wish to have access to its content have had a fair chance to consult it. There is no case for believing that papers circulated in electronic form between people whose computers happen to be connected to a common network can be said to have been published in the true sense. It would be different if publication by electronic network were physically accessible to all potential readers, possible by the payment of a fee.

Much the same is true of the input side. When career prospects often depend on publication in reputable journals, there is obviously an element of injustice in any impediment to any author's submission of what he wants to publish to what he considers the most appropriate places, but that is only a small part of the reason why it would now be premature to insist that all submissions should be electronic. It is even more important that any process that makes journals more accessible to some authors than to others skews the contribution of the scientific community to its literature, by legend its only durable product.

This is why there is persistent European resentment at the page charges levied on authors and their institutions by journals published in the United States, ironically

even when the purpose of a communication is to correct a published error. (The other side of this coin is US resentment at high subscription prices for European journals.) These crucial considerations are too often overlooked, especially by grant-making agencies anxious to encourage the new technology wherever possible. Not that there is no place for computers and their ancillaries in the publication process. Much has and will be done to improve the efficiency with which journals manage their internal affairs; in due course there may even be such improvement that page charges and subscription prices will be reduced. This is all that *Complex Systems* plans, for the time being (but it is surprising that a journal using so much technology internally should not provide optical character recognition for those submitting manuscripts not formatted in `LATEX`, threatening them with delay instead).

When the circumstances will change is anybody's guess. Much will depend on the pace at which the academic computer networks develop. At present, most laboratories in the United States have access to at least one of three packet-switched communications systems, but other countries tend to have separate national or regional networks, whose connections with the US systems are tenuous at best. And some large parts of the scientific community — the Soviet Union is the obvious example — have no connections at all. Elsewhere, there are no networks. If UNESCO were well-managed, the design of a global communications system would be an admirable goal. By default, as they say in the computer trade, it might be a job for the International Council of Scientific Unions. But until the job is done, the electronic part of the electronic journal will be strictly internal.

Meanwhile, now that Wolfram has emerged as one with a proper sense of tradition (and good sense), perhaps even willing to accept the force of the argument that authors enjoy seeing in print what they have written, there may be a case for his resurrecting another tradition, and calling his new journal after himself. It is only a century ago, after all, that Poggen-dorf was a name to conjure with in the literature of chemistry. In spite of the distinction of the editorial board he has recruited, Wolfram's new journal will have something of the same stamp.

John Maddox

Lower-mantle composition

High-pressure melting studies

Claude Herzberg

GEOCHEMICAL variations within the Earth's interior are an important guide to the history of the planet as it accreted, differentiated and cooled. This makes an understanding of the processes involved one of the most important problems of geophysics. Recent laboratory studies of the melting of samples with mantle compositions at high pressure have simulated conditions in the upper mantle and transition zone^{1,2}. Now experiments mimicking the lower mantle, with pressures up to 25 gigapascals (GPa)—250,000 atmospheres—are reported by Ito and Takahashi on page 514 of this issue³ and also by Ohtani and Sawamoto⁴.

The experiments show that upper-mantle peridotite compositions have the properties of liquids formed at or near their eutectic-like equilibria and thermal minima⁵. Their pressures of stability are yet to be calibrated, but probably span the range 10–25 GPa. The logical inference is that the upper mantle formed from the solidification of these liquids.

It would be a surprise indeed if the planetesimals that accreted to make the

Earth had compositions similar to these peridotite liquids. If instead the bulk composition of the Earth is chondritic (with a lower Mg/Si ratio), then the lower mantle must be rich in SiO₂ compared with the upper mantle, probably as a pyroxene phase such as perovskite. But is there a mechanism that could have caused this differentiation to occur?

The new experimental studies are significant because they provide a simple answer to this question. They show that the liquidus phase for peridotite and chondrite compositions is majorite from about 13 to 25 GPa, and perovskite^{1–4} at about the same or higher pressures. Both, of course, are high-pressure polymorphs of pyroxene. Because they are much more dense than peridotite liquids, they could have fractionated into the lower mantle. In effect, the lower mantle could contain a significant cumulate/residuum component⁶ of liquidus perovskite or of liquidus majorite transformed to perovskite. It is also reasonable to infer that this differentiation event happened soon after the Earth formed, when it was hot enough

for melting to have extended to great depths. This hypothesis would predict geochemical differences between the upper and lower mantles that depend on the amount of majorite and/or perovskite fractionated.

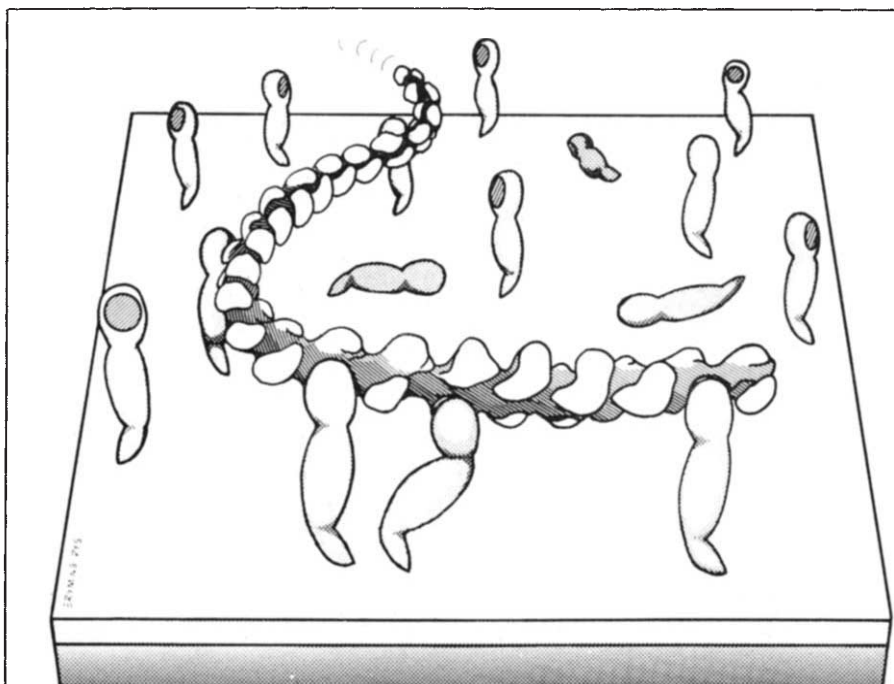
From a geochemical interpretation of their data, Ito and Takahashi decide that majorite fractionation could not have occurred, and that perovskite fractionation was not important after all. There are important difficulties, however, with their interpretation. Specifically, modified spinel eventually joins liquidus majorite as a coprecipitating phase, but is not included in the analysis of Ito and Takahashi. Problems also arise for the ratios Ca/Al and Al/Si in majorite, because they vary with bulk composition^{2,4}.

Kato, Irifune, and Ringwood⁶ also conclude that majorite or perovskite fractionation was not important, basing their judgement on major-element and rare-earth-element partitioning experiments between majorite and liquid. Again, there are problems with this interpretation. The major-element geochemistry of majorite crystallizing from the komatiite composition used in their experiments is a poor test of the hypothesis because it differs from liquidus majorites in a chondritic Earth^{2,4}. The rare-earth elements, however, should be a quintessential record of majorite or perovskite fractionation had it occurred⁶, yielding a primordial upper mantle that was high in the gadolinium to ytterbium ratio and enriched in the light rare-earth elements. Indeed, this kind of garnet signature is typical of the ancient granular-garnet peridotite nodules from South Africa. As for the present-day mantle from which mid-ocean-ridge basalts are produced, would the geochemical record of liquidus majorite or perovskite fractionation be preserved after 4 billion (4×10^9) years of partial melting?

These are all important considerations when modelling the Earth's interior. In the whole-mantle convection model, perovskite may have to become concentrated somewhere in the lower mantle because its geochemical signature is not obvious in any mantle rocks studied so far. In the two-layer convection model, silica enrichment might occur at the 670 kilometre discontinuity. Whatever the correct interpretation, the new experimental work^{1–4} from Japan demonstrates that it is difficult to avoid the conclusion that the lower mantle is enriched in silica. □

1. Takahashi, E. *J. Geophys. Res.* **91**, 9367–9382, (1986).
2. Ohtani, E. & Sawamoto, H. *Nature* **322**, 352–353, (1986).
3. Ito, E. & Takahashi, E. *Nature* **328**, 514–517 (1987).
4. Ohtani, E. & Sawamoto, H. *Geophys. Res. Lett.* **14**, 733–736 (1987).
5. Herzberg, C.T. & O'Hara, M.J. *Geophys. Res. Lett.* **12**, 541–544 (1985).
6. Kato, T., Irifune, T. & Ringwood, A.E. *Geophys. Res. Lett.* **14**, 546–549 (1987).

Claude Herzberg is in the Department of Geological Sciences, Rutgers University, New Brunswick, New Jersey 08903, USA.



MYOSIN, a two-headed molecule with a long tail, is the protein that hydrolyses ATP to produce muscle contraction. On page 536 of this issue, Toyoshima *et al.* describe a new *in vitro* motility assay for proteolytic fragments of myosin. As previously reported in the News and Views article by Clive Bagshaw (*Nature* **326**, 746; 1987), they observed movement of actin filaments over subfragment 1, the isolated myosin head, randomly arrayed on a nitrocellulose film, as depicted above. They find that the movement directed by subfragment 1 and myosin is similar, suggesting that the molecular engine of myosin is fully within the myosin head. □

Archaeology

Unsteady date of a big bang

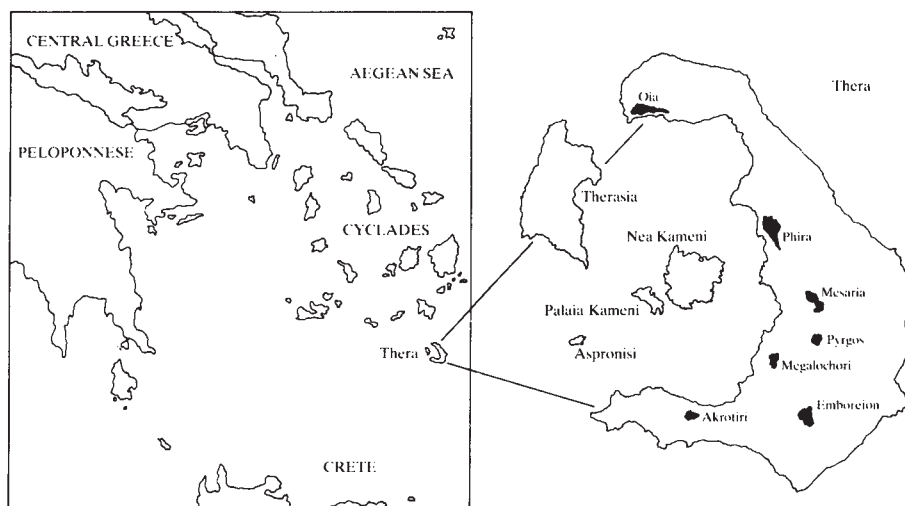
Gerald Cadogan

APPROXIMATELY 3,500 years ago, a catastrophic volcanic eruption on the Aegean island of Thera (Santorini) engulfed the town of Akrotiri (see map). There has been lively debate on several issues concerning Thera since 1967, when Spyridon Marinatos began excavating Akrotiri. A dramatic and perplexing development in this debate is reported on page 517 of this issue by C.U. Hammer *et al.*¹ who report an independent dating of the eruption, obtained by studying acidic fallout trapped in the Greenland ice core. This result, together with other recent datings, throws studies of the period into a quandary.

Akrotiri, often called 'Thera', is rightly

1500 BC (on the assumption that this happened as or immediately after the town was finally abandoned), and for the whole cultural phase 1550–1500 BC.

The ice-core date is an independent check of the archaeological date, which depends on a complex group of correlations with the calendrical chronology of pharaonic Egypt, and of radiocarbon dates from Thera. The date 1645 BC obtained by Hammer *et al.* in this issue falls within the two standard deviations (2σ) average range of the calibrated radiocarbon dates (1675–1525 BC) from short-lived material (mostly seeds). Their range is also close to Aitken's⁴ average



Map showing the position of the Thera island group, whose culture combined the traditions of the Cyclades with new ideas from Minoan Crete. The enlargement shows the location of Akrotiri and the other main archaeological sites on the island of Thera.

known as a prehistoric Pompeii. Its rich culture preserved by the volcanic fallout shows a blend of the Middle Bronze Age traditions of the Cyclades with new ideas from Minoan Crete, notably in buildings, daily paraphernalia of life and iconography. It is an important marker in the assessment of the growth of Aegean and east-Mediterranean civilizations.

The civilization belongs in archaeological terms to the Late Cycladic I and the Late Minoan IA cultural phases. The 'Minoan' system for dating the Aegean Bronze Age, with Early, Middle and Late Minoan, each with three main divisions and further subdivisions, follows that of the Old, Middle and New Kingdoms of Egypt. This Minoan model was copied for the Helladic and Cycladic cultures, which have more or less parallel sequences.

The conventional archaeological date^{2,3} for the volcano's destruction of Akrotiri is

range of 1670–1520 BC. Both Aitken and Hammer *et al.* use the new high-precision calibrations of Pearson and Stuiver⁵.

But 1645 BC does not agree with the independent date of LaMarche and Hirschboeck⁶, who dated frost damage in Californian bristlecone pines to 1628–1626 BC. These authors tentatively ascribe the damage to the effect of cold winters brought on by the release of volcanic dust into the stratosphere by the eruption of Thera. The frost-ring date falls well within 2σ , and almost within 1σ , of the radiocarbon dates of Hammer *et al.* and of Aitken, but is two decades younger than the ice-core date.

There is also disagreement between the ice-core date and the conventional archaeological date (1550–1500 BC) for Late Minoan IA/Late Cycladic I. But Betancourt⁷ has proposed recently a revised archaeological date for Late Minoan IA. Starting from the cluster of calibrated

radiocarbon dates from Thera "in the seventeenth century BC", he re-examined the archaeological correlations to produce a new Late Minoan chronology that many Aegean archaeologists will find hard to accept. His arguments depend on the stylistic dating and contextual dating of features of Minoan culture that may be correlated with Egypt. He wishes to date the phase to about 1700–1610 BC. (Note that he had not seen the new calibrations when he wrote his paper. Their effect is to make the sixteenth and seventeenth centuries BC dates equally likely.)

A principal difficulty with introducing a radically different chronology is its 'push-me-pull-you' effect. If the seventeenth century BC belongs to the Late Minoan IA phase, then the subsequent phases must be stretched out — there is no dispute about the date of the end of the Late Minoan period several centuries later — and a much reduced span must be given to the preceding Middle Minoan period in which the Minoan palaces were founded. Warren⁸ has reviewed Betancourt's arguments and suggests that Late Minoan IA started at about 1600 BC or a little earlier. This is an attractive idea in view of the considerable amount of cultural history — notably massive building programmes in Crete and on the Aegean islands like Thera — that took place in this phase and of the rather early look (still in the Middle Bronze Age tradition) of some of the pottery in the Late Minoan IA/Late Cycladic I destruction of Akrotiri.

So, which date is preferable for the eruption (and abandonment, if contemporary) of Thera? Most archaeologists would find 1645 BC or 1628–1626 BC difficult to accept, but such dates could fit the radiocarbon span. A date in the later part of the sixteenth century BC would be archaeologically acceptable, and could fit the radiocarbon span, whereas 1500 BC lies outside the radiocarbon span and is probably no longer tenable. As for the evidence for 1645 BC and 1628–1626 BC, can we be sure that there were no other causes, volcanic or non-volcanic, that could have produced the acid fallout and/or frost-ring damage? Taking everything into account, a date of the sixteenth century BC seems the most likely. □

1. Hammer, C.U., Clausen, H.B., Friedrich, W.L. & Tauber, H. *Nature* **328**, 517–519 (1987).

2. Cadogan, G. *Archaeometry* **20**, 209–214 (1978).

3. Barber, R. L. N. and MacGillivray, J. A. in *The prehistoric Cyclades. Contributions to a workshop on Cycladic chronology* (eds MacGillivray, J. A. & Barber, R. L. N.) 296–302 (Department of Classical Archaeology, Univ. of Edinburgh, 1984).

4. Aitken, M. J. in *Proc. Third Archaeometry Meeting Brit. Sch. Athens* (in the press).

5. Pearson, G.W. & Stuiver, M. *Radiocarbon* **28**, 805–838 (1986).

6. LaMarche, V.C.J. & Hirschboeck, K.K. *Nature* **307**, 121–126 (1984).

7. Betancourt, P.P. *Archaeometry* **29**, 45–49 (1987).

8. Warren, P. M. *Archaeometry* (in the press).

Gerald Cadogan is Archaeology Correspondent of the Financial Times and lives at The Old Rectory, Culworth, Banbury OX17 2AT, UK.

Stellar astronomy

The chicken came first

Virginia Trimble

THE hot interiors of massive stars produce most of the elements heavier than hydrogen and helium (metals to astronomers) that we see around us. Conversely, a star's initial small supply (0.01–4 per cent) of heavy elements is a major determinant of its observable properties and evolutionary history. Thus arises a standard sort of 'which came first' problem that now seems to have been answered.

For an aggregate of stars, the single most important property is the initial mass function (IMF), the number of stars formed per unit time as a function of mass from 0.1 to 100 or so times that of our Sun. Knowing the IMF in a group of stars enables us to predict its luminosity, colour, mass-to-light ratio, and the rates of change of these observables and of its chemical composition. Thus, for stellar populations, the chicken-and-egg question can be phrased as, which came first, a particular initial abundance of heavy elements or a particular IMF?

Two Canadian astronomers, Graeme H. Smith and Robert D. McClure of Dominion Astrophysical Observatory, may have answered that question, at least for the stars in the globular clusters of our own Galaxy (*Astrophys. J.* **316**, 206–212; 1987). The IMF came first and determined the metal abundances of the clusters through a process called self-enrichment. Self-enrichment means simply the idea that the most massive (and so shortest-lived) stars in a cluster can complete their evolution and eject their metallic products quickly enough to pollute gas that is still forming the lower-mass stars that live long enough for us to see them today.

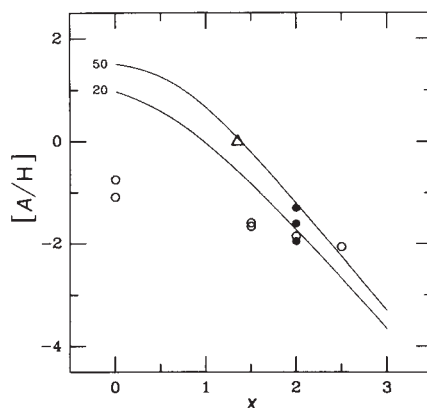
The evidence is this: most globular clusters are quite homogeneous. Thus each can be characterized by a single number representing its fractional metal content or metallicity and another number, x , the slope of the IMF, written in the form

$$N(M) dM = N_0 M^{-x} dM$$

Earlier work by McClure *et al.* (*Astrophys. J.* **307**, L49; 1986) had shown a correlation between metallicity and slope, in the direction of metal-rich clusters having smaller values of x (flatter IMFs), meaning a larger ratio of high-mass to low-mass stars. This could have meant several things. Metal-rich gas might have formed clusters with more big stars (though a best guess at formation mechanisms suggests a correlation in the opposite direction). Or, the most metal-rich clusters might have undergone the most dynamical evolution and so have preferentially lost their low-

mass stars through tidal effects (there is perhaps some evidence for this). Or, self-enrichment might have occurred, so that the clusters that began with the largest proportion of massive stars ended up with the most metals in their long-lived, low-mass stars.

Smith and McClure have now gone back and looked at the same clusters again, attempting to estimate how much metal production could be attributed to the numbers of massive stars implied by each of the IMF slopes. Their conclusion



A comparison of the observed and calculated metallicity (A/H) as a function of x , the slope of the initial mass function IMF. The calculations use stellar masses in the ranges of 0.1–20 and 0.1–50 solar masses. The observations were by R. D. McClure *et al.* $\circ$ and R. Lupton, J. E. Gunn and R. F. Griffin $\bullet$. $\triangle$ represents the solar neighbourhood parameters, $x=1.35$ and $(A/H)=0$. (From Smith, G. & McClure, R. D. *Astrophys. J.* **316**, 206–212; 1987.)

Protein toxins

Closing in on ricin action

Sjur Olsnes

RICIN, a toxic plant protein, is widely used to arm monoclonal antibodies for targeted cell killing (immunotoxins). Despite this extensive use, the mechanism of its action at the molecular level has long remained obscure. Now, work by Endo, Tsurugi and co-workers demonstrates that the A-chain of ricin is a highly specific *N*-glycosidase that removes a single adenine residue from ribosomal RNA. Another striking new observation is that cytotoxins made by bacteria involved in enteric and renal diseases are functionally and structurally related to ricin.

The general mechanism of ricin intoxication has been known since the 1970s

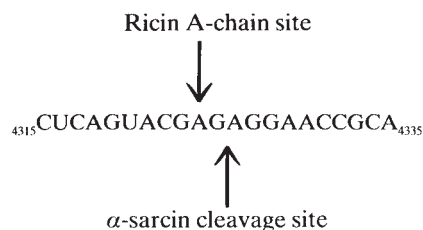
is that, for the steepest IMFs and lowest current metallicities, self-enrichment comes out just right. For the flatter IMFs there is an embarrassment of riches, and some of the metals expected from the massive stars must actually have been lost to the clusters. This is not a surprise. When many supernovae explode close together, they will set up winds that can easily expel gas from the shallow gravitational potentials of globular clusters before all the product metals can be incorporated into low-mass stars.

If the IMF determines metallicity in globular clusters, then we have to ask what determines the IMF. Smith and McClure suggest that it is the average initial density of the gas that forms a cluster (low densities yielding steep IMFs). The degree of turbulence or rotation could also be important; few globular clusters today show evidence of rotation, but they are mostly metal-rich ones.

It should be possible to extend Smith and McClure's argument by asking whether the same correlation and probable mechanism applies to the nearby dwarf elliptical galaxies. Because these are further away than the globular clusters, it will be difficult, but perhaps not impossible, to measure the slope of the IMF by counting individual stars. The average metallicities are already reasonably well known. For larger galaxies, there is no doubt that the present metal content reflects many generations of self-enrichment. And so, if metallicity is not the primary determiner of IMF, Smith and McClure's work predicts that the two should not be correlated in younger star clusters in our own and other galaxies. $\square$

Virginia Trimble is Professor of Physics at the University of California, Irvine, California 92717 and is Visiting Professor of Astronomy at the University of Maryland, College Park, Maryland 20742, USA.

(see ref. 1 for review). The toxin consists of two disulphide-linked polypeptides termed A and B. The B-chain binds to galactose-containing structures at the cell surface and facilitates entry of the A-chain into the cytosol, where it inactivates the 60S ribosomal subunits, causing disruption of protein synthesis and cell death. It was soon discovered that the toxin inactivates the 60S ribosomal subunits catalytically and that no cofactor is required, leading to the suspicion that ricin has endonuclease activity, like two other ribosome-inactivating proteins, colicin E3 and the fungal toxin α -sarcin. But despite many attempts, nucleolytic activity of



Ricin A-chain removes adenine from A₄₃₂₄ located close to the α -sarcin cleavage site in a highly conserved region near the 3' end of 28S RNA. Subscript numbers indicate the position of the bases in the sequence, counted from the 5' end. The total length of 28S RNA from rat liver is 4,813 nucleotides.

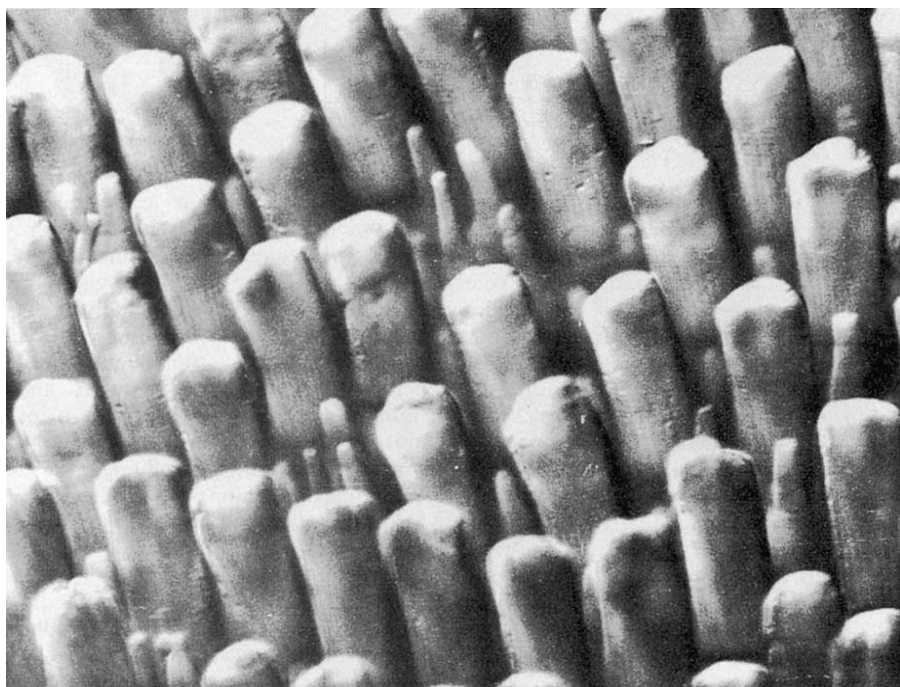
ricin A-chain was never demonstrated convincingly.

Endo and colleagues^{2,3} also searched in vain for an RNase activity associated with ricin. But, unlike previous workers, they noticed an almost undetectable reduction in the electrophoretic migration rate of 28S RNA from ricin-treated ribosomes. A 550-nucleotide fragment derived from the 3' end of the 28S RNA migrates slightly more slowly than the corresponding fragment from untreated ribosomes. Their subsequent analysis of the 550-nucleotide fragment allowed them to elucidate the mechanism of enzymatic action of the ricin A-chain.

In 28S from ricin-treated ribosomes, adenine is missing from position A₄₃₂₄ although the phosphoribose backbone is left intact (see figure). This renders the bond between G₄₃₂₃ and A₄₃₂₄ nuclease-resistant, whereas the phosphodiester bonds on both sides of A₄₃₂₄ become hypersensitive to cleavage by alkali and by anilin at low pH. Interestingly, A₄₃₂₄ lies in a region that is strongly conserved between species.

Endo *et al.* in their new work also show that ricin removes close to one adenine residue per ribosome, indicating that A₄₃₂₄ is the sole site of action. Intact ribosomes and 60S subunits are the most sensitive substrates, but native 28S RNA is also modified, although at a reduced rate. On the other hand, denatured 28S RNA and the 550-nucleotide fragment are resistant. Apparently, ricin A-chain recognizes the three-dimensional structure around A₄₃₂₄. Consistent with this observation, *Escherichia coli* ribosomes, which also contain the conserved sequence, are resistant to the enzymatic activity of ricin A-chain², and the toxin does not inhibit their action.

Several plant toxins have a similar structure and action to ricin. Endo *et al.* found that two of these, abrin and modeccin, remove adenine from A₄₃₂₄. They also found the same enzymatic activity in pokeweed antiviral protein, a member of a large group of A-chain-like proteins. These proteins inactivate ribosomes in cell-free systems, but because they lack a B-chain, are not toxic to intact cells. A-chain-like proteins are present in



Rods of the vertebrate retina, seen here dwarfing their neighbouring cones, register the capture of photons by changing their membrane voltage. This signal is passed to higher-order visual neurons through a chemical synapse, and to adjacent photoreceptors through electrical synapses. On page 522 of this issue, D. Attwell, S. Borges, S. Wu and M. Wilson report the surprising finding that only signals in the first one-fifth of the voltage response range of the rods are actually transmitted across the chemical synapse, and larger signals are clipped. The authors suggest that the electrical synapses between photoreceptors help to avoid this clipping, by allowing light to produce a small signal in many photoreceptors, each of which has its chemical synapse operating at a high gain. The average diameter of the top of a rod is 11.5 μ m. □

high amounts in a vast number of plants, including such familiar species as wheat and carnation¹, but their function is not understood.

Although the cytotoxin from the bacterium *Shigella dysenteriae* was known to inactivate catalytically the 60S ribosomal subunit⁵, it came as a surprise that this bacterial toxin acts like ricin in that it removes adenine from A₄₃₂₄ (ref. 3). Closely related toxins (Shiga-like toxins) are produced by several enteropathogenic bacteria and one such toxin, which has recently been sequenced⁶, shows marked homologies to ricin A-chain. Possibly, therefore, the bacteria have picked up a common plant gene for their own malevolent purposes.

A major problem that remains to be solved is how ricin crosses the cell membrane to reach its targets in the cytosol. Endocytic uptake must be involved, but transfer to the cytosol seems to come from a compartment distal to endosomes. Ultrastructural and biochemical data have focused attention on the trans-Golgi network as the possible transfer site^{7,8}. Ingenious experiments recently published by Youle and Colombatti⁹ support this possibility. These authors tested the ricin sensitivity of hybridoma cell lines that produce anti-ricin antibodies. On their way to the surface the antibodies, like other export

proteins, pass through the trans-Golgi network where they can meet endocytosed ricin. These hybridoma cells are highly resistant to ricin, and appropriate controls exclude the possibility that antibodies in the medium or at the cell surface cause the resistance.

Considerable progress has also been made in the elucidation of ricin structure. The toxin had earlier been sequenced and cloned, but now the three-dimensional structure has been solved at 2.8-Å resolution¹⁰ and data at 2.4-Å resolution are expected soon. The penetration process can therefore now be probed by site-directed mutagenesis of selected regions of the toxin. □

1. Olsnes, S. & Pihl, A. in *The Molecular Actions of Toxins and Viruses* (eds Cohen, P. & van Heyningen, S.) 52-105 (Elsevier Biomedical, New York, 1982).
2. Endo, Y. *et al.* *J. biol. Chem.* **262**, 5908-5912 (1987).
3. Endo, Y. & Tsurugi, K. *J. biol. Chem.* **262**, 8128-8130 (1987).
4. Stirpe, F. & Barbieri, L. *FEBS Lett.* **195**, 1-8 (1985).
5. Reisbig, R., Olsnes, S. & Eiklid, K. *J. biol. Chem.* **256**, 8739-8744 (1981).
6. Calderwood, S.B. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **84**, 5398-5403 (1987).
7. van Deurs, B. *et al.* *J. Cell Biol.* **102**, 37-47 (1986).
8. Sandvig, K. *et al.* *Cancer Res.* **46**, 6418-6422 (1986).
9. Youle, R.J. & Colombatti, M. *J. biol. Chem.* **262**, 4676-4682 (1987).
10. Montfort, W. *et al.* *J. biol. Chem.* **262**, 5398-5403 (1987).

Sjur Olsnes is at the Institute for Cancer Research at the Norwegian Radium Hospital, Montebello, Oslo 3, Norway.

Nuclear physics

Beyond the neutron drip line

P.G. Hansen

APPROXIMATELY 288 stable or near-stable nuclear species, characterized by the number of neutrons, N , and the number of protons, Z , occur in nature. The total number of nuclear systems (N, Z), radioactive but stable against prompt particle emission, is predicted to be about 6,000, but only about one-third of these have been observed. As shown in the figure, which does not include heavy and super-heavy elements, the neutron-rich nuclei, many of which have yet to be identified, present a major challenge. For most elements, we simply do not know of reactions that will allow us to probe the region of the neutron drip line (see figure legend). Within the past few years there has, however, been remarkable progress in the techniques for studying the lightest elements, as exemplified by the recent work of Seth and co-workers. In a recent paper¹ they describe a spectroscopic study of the nucleus ${}^9\text{He}$, which has five more neutrons than the usual ${}^4\text{He}$ nucleus, taking it beyond the neutron drip line.

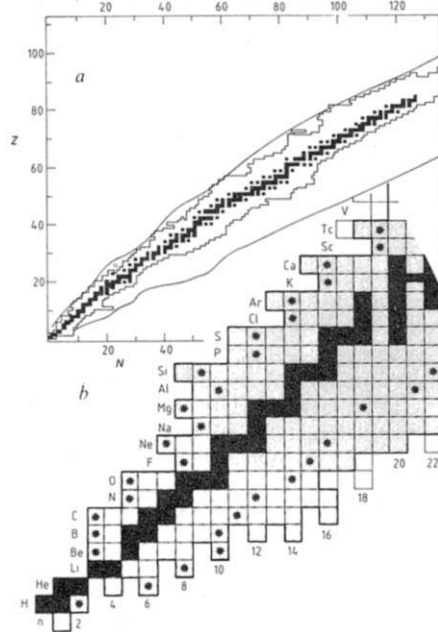
Working at the Los Alamos Meson Physics Facility, Seth and co-workers used a special pion spectrometer (EPICS) to observe double-charge-exchange reactions of the type ${}^A\text{Be}(\pi^-, \pi^+){}^A\text{He}$, in which a negatively charged pion impinging on the stable beryllium nucleus is converted into a positive pion, and two protons in the nucleus become neutrons, to give the helium nucleus. The spectroscopy involves determining the intensity of the scattered pions as a function of their loss of energy (having allowed for the recoil of the target nucleus). A resonance in the spectrum corresponds to a bound or quasi-stationary state of the product nucleus.

Early results reported at the 1981 Helsingør conference² showed that the spectrum is dominated by a broad distribution, representing all the possible states of the non-resonant break-up products, and also a peak attributed to the ${}^9\text{He}$ ground state. The new experiment, with improved resolution and statistics, confirms the analysis and also reveals additional peaks representing excited states in ${}^9\text{He}$. The ground state is unbound by 1.13 ± 0.10 MeV with respect to neutron emission. Because the resonances have no measurable intrinsic width, it can be deduced from Heisenberg's uncertainty principle that the ${}^9\text{He}$ states live for at least 10^{-20} s, so that this nucleus survives for at least several periods of its intrinsic motion before breaking up.

It is interesting that the measured ground-state masses of the neutron-rich helium nuclei turn out to be lower, which

is to say that the systems are more bound than semi-empirical estimates ('Garvey-Kelson relations') based on measured masses from near stability suggest. This could mean that these nuclei adjust to the neutron excess by assuming a structure that is different from the one encountered near stability. If so, theoretical extrapolations to the neutron drip line are much more uncertain than normally assumed.

The double-charge-exchange reaction is of limited value for studying the drip line for heavier systems. Seth *et al.*¹ point out that a search for the ${}^{10}\text{He}$ ground state would require a target of long-lived, radioactive ${}^{10}\text{Be}$, not easily available in usable quantities. The heaviest unbound system that can be reached from a natural target is ${}^{13}\text{Be}$, so the prospects for progress in this direction are limited.



The nuclear chart. Vertical scale, increasing nuclear charge, Z , on which the chemical identity of the nucleus depends; horizontal scale, neutron number, N . Black squares, stable isotopes. In *a*, the zig-zag lines are the experimental limits of known isotopes; the two smooth lines indicate the estimated limits towards prompt instability, the proton (upper) and neutron (lower) 'drip lines'. *b* shows the status for the light nuclei. The heavy line surrounds the (shaded) area of isotopes detected experimentally with half lives longer than 10^{-7} s; external squares in thin lines, predicted but as yet unobserved isotopes. Stars, heaviest and lightest isotopes for which the half life is known experimentally. All isotopes of all elements up to nitrogen now seem to have been detected. Note also that the proton drip line has been reached for many medium-weight and heavy elements and, apparently, for all light elements up to scandium, where $Z = 21$.

Fortunately, fragments from heavy-ion reactions at intermediate energy can be identified in flight, making possible studies of all bound isotopes of the lighter elements^{3,4}. In the past two years, this method has reached a new level of perfection in a series of experiments⁵⁻⁸ carried out at GANIL, the French heavy-ion laboratory in Caen, where energies of 30–100 MeV per nucleon are available. With the observation^{5,6} of the nuclides ${}^{22}\text{C}$ and ${}^{23}\text{N}$ there, all particle-stable isotopes of the seven elements up to nitrogen have been produced in the laboratory.

The essential tool in the GANIL experiments is the magnetic spectrometer LISE. Two dipole magnets select the mass-to-charge ratio A/Z , and a degrader, placed between the magnets, causes an energy loss proportional to Z^2 . The net effect is a complete separation by mass and element, allowing the study^{7,8} of radiations from the stopped radioactive fragments. The flight time through the apparatus, about 10^{-7} s, limits the technique to nuclei with half lives longer than this value.

An interesting hint of an open problem in nuclear physics emerges from the pattern of particle-stable nuclei shown in the lower part of the figure, which combines information from many sources. The preponderance of even- N isotopes near the neutron drip line suggests that neutron pairing is essential for nuclear stability in this region. Pairing in nuclear parlance denotes a structure in which the neutrons (or protons) are coupled in spin-zero pairs, where the two nucleons have the same orbital quantum numbers, and where a unique combination of pairs gives rise to a favoured, low-lying spin-zero ground state in even-even nuclei. For example, the helium isotopes are stable for masses $A = 4, 6, 8$, and unstable for $A = 5, 7, 9$. The seven elements up to nitrogen show the same tendency: in each case the heaviest bound isotope has even $N (=2n)$ and in all but two cases the isotope with $N=2n-1$ is unbound. In two cases (helium and boron), the isotope with $N=2n-3$ is also unbound.

Superficially, the importance of pairing agrees well with the conventional wisdom that even- N nuclei are more strongly bound by $12A^{-1/2}$ MeV than odd- N nuclei (for even Z). The pairing strength, however, decreases strongly with increasing neutron excess, as expressed by three approximately equivalent, semi-empirical relations⁹⁻¹¹ including an isospin- (or charge-) symmetry term. Extrapolations to the neutron drip line of these relations predict that pairing vanishes there. This is manifestly not the case. Neither the old treatment of pairing, which works well near the bottom of the stability valley, nor the new, improved approximation seems to hold near the drip line. The upshot is that experiments, in particular very accurate determinations of mass, will again

have the last word, and our ideas about stability and structure of very neutron-rich nuclei will change in the next few years. □

1. Seth, K.K. *et al.* *Phys. Rev. Lett.* **58**, 1930–1933 (1987).
2. Seth, K.K. *Proc. 4th Int. Conf. Nuclei Far from Stability*, CERN 81–09, 655–663 (1981).
3. Artukh, A.G. *et al.* *Nucl. Phys.* **A176**, 284–288 (1971).
4. Symons, T.J.M. *et al.* *Phys. Rev. Lett.* **42**, 40–43 (1979).
5. Langevin, M. *et al.* *Phys. Lett.* **150B**, 71–74 (1985).
6. Pougheon, F. *et al.* *Europhys. Lett.* **2**, 505–509 (1986).

7. Dufour, J.P. *et al.* *Z. Phys.* **A324**, 487–488 (1986).
8. Dufour, J.P. *et al.* *Nucl. Instr. Meth.* **A248**, 267–281 (1986).
9. Vogel, P., Jonson, B. & Hansen, P.G. *Phys. Lett.* **139B**, 227–230 (1984).
10. Jensen, A.S., Hansen, P.G. & Jonson, B. *Nucl. Phys.* **A431**, 393–418 (1984).
11. Madland, D.G. & Nix, J.R. *Nucl. Phys.* (in the press).

P.G. Hansen is at the Institute of Physics, University of Aarhus, DK 8000 Aarhus C, Denmark.

Biomechanics

Wallabies vibrate to breathe

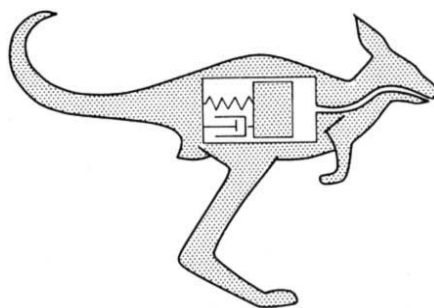
R. McNeill Alexander

WALLABIES take one breath per hop throughout their range of hopping speeds. Baudinette and his colleagues at Flinders University¹ demonstrate this result in experiments in which the wallabies hopped on a motorized treadmill. They suggest that the respiratory system is a mechanical oscillator and breathing is largely passive, driven by hopping movements.

Similar suggestions have been made for horses, dogs and jackrabbits, which take one breath per stride when they gallop². But wallabies are more convenient subjects because they travel on just two legs which move in phase with each other (see photograph), making it easier to interpret the data. Galloping quadrupeds move each leg out of phase with all the rest and have correspondingly complicated patterns of acceleration and deceleration³. People also coordinate breathing movements with their running movements, but usually take two strides per breath.

The hypothetical oscillator of wallabies is shown in the sketch. The movable mass inside the animal represents the abdominal viscera, which move posteriorly as the lungs fill and anteriorly as they empty. It is mounted on a spring, which represents the elastic properties of the diaphragm and other tissues, and a dashpot, which represents the damping effects of tissue and air viscosity. Diagrams like this appear in engineering textbooks, representing accelerometers and seismometers⁴. If the spring were stiff enough to make the natural frequency of the system much higher than the hopping frequency, the viscera would behave like an accelerometer. They would make only

small movements relative to the skeleton, and their displacement at any instant would be proportional to the body's acceleration. The lungs would be fullest at the instant of peak forward acceleration. However, if the spring were compliant enough to make the natural frequency of



the viscera much lower than the hopping frequency, they would act like a seismometer. They would travel at almost constant velocity (equal to the animal's mean velocity) while the animal accelerated and decelerated. The lungs would enlarge while the animal was travelling faster than the mean and get smaller while it was slower, and they would be fullest when it had peak deceleration.

The remaining possibility is that the natural frequency of the viscera is about equal to the stride frequency. If it were tuned precisely to the stride frequency, and if the animal's pattern of acceleration and deceleration were sinusoidal, the displacements of the viscera relative to the skeleton would be in phase with the animal's fluctuating velocity. The lungs would be fullest when the animal was travelling fastest (when its acceleration

would, of course, be zero). The pattern of acceleration is unfortunately not sinusoidal, so the situation is more complicated. The animal decelerates and re-accelerates while its feet are on the ground and travels with a constant horizontal component of velocity while they are off the ground⁵. (I am ignoring a further complication: the system may be affected by vertical accelerations as well as horizontal ones.)

For the system to function as an accelerometer, the diaphragm and body wall would have to be so stiff that the amplitude of the breathing movements would be very small. For it to function as a seismometer, they would have to be so compliant that the animal would be apt to leave its guts behind in a prolonged acceleration. Neither possibility seems attractive, and neither would give the observed phase relationship between breathing and hopping. Baudinette *et al.* in their new work show¹ the lungs are emptiest while the feet are on the ground, and fullest just before the wallaby reaches the highest point in its hop.

The tuned-oscillator possibility seems to predict a reasonably realistic phase relationship and (with different degrees of damping) can give almost any amplitude of breathing movements. It seems the most likely possibility, and there is an encouraging piece of further evidence. In a study of panting, Crawford⁶ took X-ray pictures of freshly-dead dogs tilted head up and tail up. He observed how far the viscera moved between the two positions, and calculated that their natural frequency of vibration would have been about 4 Hz. This is less than the frequency of panting, which involves a different mode of vibration, but not very different from the stride frequency of about 3.2 Hz that the dogs would probably have used when running⁷.

Wallabies increase their frequency of stride very little, if at all, as they increase speed from 2 to 9 m s⁻¹ (ref. 1): they go faster by taking longer strides. Similarly, quadrupedal mammals keep their stride frequencies almost constant over their whole range of galloping speeds⁷. The concept of the respiratory system as a tuned oscillator offers a possible explanation: it would work well at the frequency it was tuned to, but not at other stride frequencies. □



1. Baudinette, R.V., Gannon, B.J., Runciman, W.B., Wells, S. & Love, J.B. *J. exp. Biol.* **129**, 251–263 (1987).
2. Bramble, D.M. & Carrier, D.R. *Science* **219**, 251–259 (1983).
3. Cavagna, G.A. *et al.* *Am. J. Physiol.* **233**, R243–R261 (1977).
4. Thomson, W.T. *Theory of Vibration with Applications* 2nd edn (Allen & Unwin, London, 1983).
5. Alexander, R. McN. & Vernon, A. J. *Zool., Lond.* **177**, 265–303 (1975).
6. Crawford, E.C. *J. appl. Physiol.* **17**, 249–251 (1962).
7. Heglund, N., McMahon, T.A. & Taylor, C.R. *Science* **186**, 1112–1113 (1974).
8. Bramble, D.M. *Proc. N. Am. Congr. Biomech.* **1**, 3–4 (1986).

R. McNeill Alexander is Professor of Zoology at the University of Leeds, Leeds LS2 9JT, UK.

Biological productivity

Human use of world resources

Jared M. Diamond

THERE is a continuing controversy over the extent of human appropriation of the world's biological productivity. On the one hand, some economists¹ and representatives of power companies² argue that human impact is too modest to pose imminent biological limitations on our population growth, nor were such limits foreseen in a recent well-publicized report on population growth and economic development³. On the other hand, some ecologists estimate that we are already appropriating nearly 40 % of potential net primary productivity on land^{4,5}, but this estimate has been denounced by a power-company representative as "unbelievable" and "exaggerated by several orders of magnitude"².

The estimate thus denounced was by Vitousek, Ehrlich, Ehrlich and Matson⁴, four ecologists at Stanford University and NASA/Ames Research Center, who analysed the massive international database on global productivity, carbon storage, agriculture, forestry, land use and land conversion. The table summarizes their analysis, which proceeded in four stages.

First, annual net primary production (NPP) is defined as the energy (mostly solar) fixed biologically minus the respiration of the primary producer species (mostly plants) that effect this biological fixation. NPP, measured in petagrams of organic matter (1 Pg = 10¹⁵ g) per year, is the energy left over for all consumer and decomposer species, including humans. To estimate the global NPP, the Earth's surface area is classified into biomes, productivity measurements are taken for each biome, and the area and productivity are multiplied and summed over the biomes. The resulting estimate is a global NPP of about 225 Pg (see table).

Second, the proportion of this NPP directly used by humans for food, and of wood used for timber, paper, and firewood is estimated as 7.2 Pg, or 3 % of the Earth's total NPP (see table).

Third, human acquisition of food and wood diverts far more biological productivity than is used. For example, on average less than 10 % of the total organic matter fixed on agricultural land is edible. When a natural forest or a tree plantation is harvested, slightly less than 50 % of its accumulated NPP becomes merchantable timber. When forests are cleared for shifting cultivation or permanent use, most of the organic matter felled is burnt or decomposes. The organic matter produced on other converted areas, such as lawns and golf courses, is not used. The total NPP that humans command is 42.6 Pg, of

which 7.2 Pg are used directly and 35.4 Pg diverted. This represents 19% of the Earth's total NPP: 31% of that on land, 2 % of that in the sea.

Finally, it is not the case that humans leave global NPP unchanged and merely divert some of it. Instead, we permanently modify the global NPP, because the habitats that we create intentionally (asphalt, pastures and croplands) and unintentionally (by desertification) do not have the same NPP as the natural habitats from which they were converted. These reductions in NPP (by 17.5 Pg) are in addition to the 42.6 Pg of the remaining NPP that we command. Hence our total appropriation is 17.5 + 42.6 = 60.1 Pg (see table) compared with a global NPP of 224.5. This appropriation falls far more heavily on the terrestrial NPP (17.5 + 40.6 = 58.1, out of a potential terrestrial NPP of 132.1 + 17.5 = 149.6, or 39 %) than on marine NPP (2.0/91.6 = 2 %). That leaves 61 % of the land's potential NPP for all other terrestrial consumer species. Our proportion of terrestrial NPP represents probably the largest appropriation of energy by one species and its servant species since life colonized the land.

Obviously, all the numbers involved in these calculations are being continually refined. But several conclusions are equally obvious. Our growing impact, whether somewhat more or less than 39%, is already large. The fraction of terrestrial productivity left to sustain all other species has already shrunk significantly. Even if we appropriate the entire

global NPP and exterminate all species except ourselves and our servant species, we will soon reach biological limits to our population growth, given our present doubling time of 41 years and our current mix of extraction techniques.

The estimate I have outlined above was criticized in a letter to *Science* by B. L. Godfriaux of the Public Service Electric and Gas Company, Newark, New Jersey², who based his discussion on a summary of the Vitousek *et al.* paper⁴ rather than on the paper itself. According to Godfriaux, research at his power company indicates that production of tomatoes in greenhouses is now approaching 300,000 pounds per acre. Given assumptions about the calorie content of tomatoes, calorie requirements of humans (taken as equivalent to 16 pounds of tomatoes per day), size of the US population and tillable area of land in the US, Godfriaux concludes that only about 0.56 % of US tillable lands, if devoted to greenhouse monoculture of tomatoes, would suffice to supply US human calorie requirements, and hence that the estimate of Vitousek *et al.* must be several orders of magnitude too high.

But Godfriaux does not address the question of the fraction of global NPP actually appropriated by humans⁵. In addition, the reasons why one cannot extrapolate from intensive production of tomatoes in a greenhouse to all the needs of the world's human population are many and obvious: the greenhouse requires inputs such as water, fertilizer, artificial light, heat and building materials; the high-calorie yields of tomatoes in greenhouses cannot even be approached by tomatoes grown outside greenhouses or by most other crops grown inside greenhouses: people require not only calories but also other nutrients such as protein

Human appropriation of the world's net primary productivity (NPP, in units of 10 g organic matter per year).

World NPP		NPP directly used by humans and domestic animals	
Terrestrial	132.1	Plants eaten by humans	0.8
Fresh water	0.8	Plants eaten by domestic animals	2.2
Marine	91.6	Fish eaten by humans and domestic animals	2.0
Total	224.5	Wood used for timber or paper	1.2
		Wood used for firewood	1.0
		Sub-total	7.2
NPP used or diverted by humans		NPP used, diverted or permanently reduced by humans	
Cropland	15.0	NPP used or diverted	42.6
Converted pastures	9.8	NPP reduced by conversion:	
Consumed by livestock or burnt in natural pastures	1.8	to cropland	9.0
Human-occupied areas	0.4	to pasture	1.4
Wood used for timber, paper or firewood	2.2	to desert	4.5
Forest destruction without use:		to human-occupied areas and roads	2.6
during harvesting of forest	1.3	Total	60.1
during harvesting of tree plantations	1.6		
during clearing for shifting cultivation	5.9		
during clearing for permanent use	2.6		
Fish used for food	2.0		
Sub-total	42.6		

and vitamins from foods whose growth efficiency is far below that of tomatoes (for example, domestic animals); and people require things besides food (for example, inedible materials, energy and absorption of wastes).

This controversy about a particular ecological question raises a broader problem. Ecological information is important in evaluating many other questions of public policy as well as human population growth and use of resources. Scientists are becoming increasingly troubled that governments and the public often ignore them. Yet the scientific community itself often places a low value on ecology and population biology. The recent report³ that minimizes biological limitations on

economic development was not issued by industry but by a committee of the National Research Council of the National Academy of Sciences including economists but no ecologists or population biologists. Scientists can hardly expect the public to listen to them when their own professional organizations fail to do so. □

1. Simon, J.L. *New Scient.* **232**, 60-63 (1986).
2. Godfriaux, B.L. *Science* **235**, 15 (1987).
3. National Research Council *Population Growth and Economic Development: Policy Questions* (National Academy, Washington DC, 1986).
4. Vitousek, P.M. *et al. BioScience* **36**, 368-373 (1986).
5. Vitousek, P.M. *et al. Science* **235**, 730 (1987).
6. Lewin, R. *Science* **234**, 14-15 (1986).

Jared M. Diamond is Professor of Physiology at the University of California Medical School, Los Angeles, California 90024, USA.

Meteorology

Interactions of wind and waves

Gerbrand Komen

RESEARCH on ocean waves is evolving rapidly: for the first time an understanding of physical processes responsible for the generation of ocean waves has reached a level at which it can be (and is) used to infer the large-scale response of the ocean surface to wind forcing. At the same time application of ocean-wave modelling is going beyond traditional objectives (such as the routing of shipping) as it is becoming increasingly clear that good global knowledge of the sea state can help in the understanding weather and climate. This can work in two ways. First, measurements of wave heights can be used to modify estimates of wind speed. Second, the effect of wind-wave interactions on the boundary layer over the sea can be dramatic, and this in turn has an impact on weather patterns. These developments were the topic of a recently held workshop* of the international WAM collaboration.

Traditionally, numerical wave-prediction models are based on a mixture of theory, empiricism and *ad hoc* assumptions. But wave growth is the result of a rather delicate balance between input of energy from the atmosphere, dissipation to the underlying ocean and hydrodynamic nonlinear redistribution of energy between different wave components. The usual distinction between wind, sea and swell turns out to be highly artificial and arbitrary. As a result, models were not very accurate in complex or extreme situations, unless they had been tuned to reproduce observations under particular conditions. (See, for example, Robert Long's News and Views article in *Nature* **313**, 182-183; 1985.)

In 1984, on the suggestion of Klaus

Hasselmann (Max-Planck Institut für Meteorologie, Hamburg), an international group of wave researchers (40 people from 10 countries), not satisfied with this state of affairs, started a collaboration to develop and implement a model based on best-known expressions for input, dissipation and nonlinear transfer. They have now produced a model (WAM, for wave modelling) which has run successfully on several large computers, notably on the Cray-XMP/48 of the European Centre for Medium Range Weather Forecasting (ECMWF) in Reading, United Kingdom.

The original prototype of the model has now grown into a flexible system of programs that includes all the necessary pre- and post-processing software. The model has been tested in many hindcasts: six storms on the north-west European shelf; a comprehensive North Sea study; three hurricanes in the Gulf of Mexico; several storms in the Mediterranean; and an extended global hindcast. The results compare favourably with observations, although the model may be somewhat overpredicting in the early stage of wind sea development, when waves have just begun to grow and are still relatively low. This is being fixed now, but it suggests that our ideas about white cap coverage in these early stages of development need revision. Also, the model has been run in real time, to make actual forecasts, using wind forecasts of ECMWF as input. After tests on a regional version, the global model was run for the first time on 7 March. Since then, a wave analysis and a five-day forecast have been made each day, producing detailed global wave charts.

Other work has focused on more general wave-dynamical problems. One

group is reanalysing all existing observations of fetch-limited growth. They find for example, that air-sea temperature differences are an important factor in wave growth. The effect of changing wind direction on ocean waves can be drastic, a tragically illustrated by the Fastnet sailing race disaster in 1979. The calculations are complicated by nonlinear coupling between waves moving in different directions, but are possible now using the efficient codes of the WAM model. Also it can now be shown how in shallow water, long waves become dissipated while short waves continue to propagate.

Last, but not least, the group addresses the data-assimilation problem. It is here that the relationship to weather prediction is most clear. The general idea of data assimilation is to improve the initial model fields (resulting from earlier forecasts) by incorporating observations, with the hope of improving subsequent predictions. In weather forecasting this is common practice, as forecasting skill is known to rely heavily on having high-quality initial fields. Assimilation of data on ocean waves is now becoming possible because of the development of remote sensing especially with the use of satellite (Geosat, ERS-1) that can measure wave heights to an accuracy of 0.5 metres. These data, together with ripple-density measurements (also from satellites), can be used to obtain accurate estimates of wind speeds over the sea.

As it happens, ocean-wave prediction is intimately related to wind prediction. If computed wave height has to be modified the computed winds may have to be changed as well. Therefore, the problem of data assimilation in numerical wave and atmospheric models is coupled. In addition, the state of the sea influences the atmospheric boundary layer and the momentum flux to the ocean; it also affects satellite observations of wind above the sea. Ultimately, it is to be hoped that real-time data assimilation of all observations in combined wave and atmosphere models will be available. This would also ensure that best use is made of satellites for climate studies. Only by combining measurements and models can accurate global-stress fields, essential in climate modelling, be compiled.

The effectiveness of the WAM group may be related in part to the way in which it arose. Its work evolved in an iterative way as a result of a series of meetings in which individual plans and group tasks were presented alternately. In this way different interests and abilities, personal and institutional, were merged without any strong power structure, but with great success.

Gerbrand Komen is in the Department of Oceanography of the Royal Netherlands Meteorological Institute, PO Box 201, 3730 A De Bilt, The Netherlands.

* The WAM group met at Woods Hole Oceanographic Institution, 2-5 May 1987.

Pterodactyl habits — real and radio-controlled

SIR—Following in part new work by others^{1,2}, Unwin³ has suggested that the pterosaur's ability to move on the ground was very limited. This contradicts arguments that pterosaurs were terrestrial, erect-limbed bipeds⁴. Pterosaur anatomy and lifestyles show a diverse pattern of running and climbing adaptations.

It has recently been shown that in some pterosaurs the lower hip elements were splayed apart, and that the acetabulum faced to the side or a little up⁵. Unwin believes this means that they had sprawling, "clumsy" legs. But sprawling lizards and certain bats are quite agile, and erect birds also have widely separated lower hip elements and side-facing hip sockets. The sharply inturned, bird-like femoral head of the early pterosaur *Dimorphodon* worked in an erect plane, with the knee only moderately bowed out. In more advanced pterosaurs less offset, more spherical femoral heads indicate that they had reverted to more supple, semi-erect legs. Trackway gauge probably remained narrow because the knee was sharply flexed postero-medially. It is emphasized that the pterosaur's avian-style knee, shank and ankle were too stiff to follow the twisting motions of a truly sprawling gait. Nor did any pterosaur have reduced swift- or frigate bird-like legs that would have made them incompetent on the ground.

Quetzalcoatlus firmly establishes that pterosaurs were fast and agile on the ground. This giant lived on a flat floodplain 400 km from the palaeocoastline⁶. Restorations I helped draw up for a scale flapping flight model⁶ indicate that *Quetzalcoatlus* was 10 feet tall and massed perhaps 250 pounds; too large to roost in trees, pursue aerial prey, or to snap up fish from its habitat's limited watercourses while on the wing. Carcass scavenging was not possible either because the crane-like beak was much too weak, and even Marabou storks use their much stouter beaks only to pick up scraps dropped by vultures⁷. Instead, *Quetzalcoatlus* probably picked up fish, frogs, turtles and so on from the shallows. On windless days it had to

take off with an $\sim 30 \text{ km h}^{-1}$ run. The dense bristle-toothed jaws of smaller filter feeders such as *Pterodaustro* could also only be swished through the water as they waded.

As for how pterosaurs walked, flexing the elbow 90° downwards allowed their arms to work in a narrow gauge mode. Long-tailed rhamphorhynchoids could have been both two and four legged, except for *Dimorphodon* whose upper arm and hand were too short relative to the hindlimb for quadrupedalism. I agree with Pennycuik² that the tailless pterydactyls lacked the adaptations for full bipedalism found in tailless birds — they were more quadrupedal.

Impressions show that the wing membrane's trailing edge attached to the body⁴, or to the thigh¹, although these may be partly folded membranes overlapping the thighs. The rhamphorhynchoid's long, rudder-tipped tail acted as an auxiliary control surface. Tailless pterodactyls may have evolved their more supple legs in order to manipulate a hip-to-ankle membrane. Folded for streamlining during cruising flight, it could have deployed into an inverted-V shaped tail surface for landings, for instance. Traces of such a uropatigium appear to be preserved in two *Pterodactylus* specimens^{1,8}, despite Wellnhofer's arguments to the contrary.

As shown by squirrels, cats, and many birds, running and climbing are not mutually exclusive adaptations. Small pterosaurs were probably good at both, and the long, divergent outer toe of rhamphorhynchoids may have enhanced the grasping ability of their feet. However, the rhamphorhynchoid's long, ossified-rod stiffened tails could not fold tightly enough to be out of harm's way if their owners hung from branches by their hindfeet³.

GREGORY S. PAUL

3109 N. Calvert Street,
Baltimore, Maryland 21218, USA

1. Wellnhofer, P. A. *Naturhist. Mus. Wien* **88**, 149–162 (1987).
2. Pennycuik, C. J. *Mem. Calif. Acad. Sci.* **8**, 83–98 (1986).
3. Unwin, D. M. *Nature* **527**, 13–14 (1987).
4. Padian, K. *Paleobiology* **9**, 218–239 (1983).
5. Lawson, D. A. *Science* **187**, 947–948.
6. MacCready, P. *Engineer. Sci.* **XLIX** (2), 18–24 (1985).
7. Houston D. C. in *Serengeti; Dynamics of an Ecosystem* 263–286 (University of Chicago Press, 1979).
8. Broili, F. *Sitzungber. Bayer. Akad. Wiss. math.-naturwiss.*, 139–154 (1938).

Of fast teeth and big heads

SIR—Beynon and Wood interpret their important data on early hominid molar growth¹ to suggest that hominid growth two million years ago was 'more rapid' than that for modern teeth: "The absolutely shorter growing period of these larger molar crowns suggests that there was strong selection to form teeth quickly and have them erupted into the mouth early in development." There is a confounding problem of anthropocentrism here — and a common fallacy of adaptationist thinking in cases of pleiotropy.

Comparisons of the duration of life span phases in primates suggest that developmental rates have been slowed between monkey and ape, and slowed again between ape and modern humans; it is probably not a matter of 'why they were fast' but of 'why we are so slow'.

For our slowing, a proximate cause is not hard to find, given a tendency toward juvenilization (besides the flatter faces of later adult hominids, the progressively-reduced crown surface area of mandibular cheek teeth² is also consistent with retention of juvenile features). Precocious puberty does make heads relatively larger, compared to pelvic outlets, causing problems in subsequent generations. Given that a fetus with a big head kills not only itself but its mother and thereby other siblings, it is hard to imagine a more serious 'genetic disease' than big heads; the resulting selection against average and faster-than-average developmental rate variants must have been severe, with only slower-than-average variants surviving to raise children³.

Although the selection pressures on slowing might have been solely associated with slowed fetal development to hold down head size until the end of standard gestation, slower somatic development probably carried over to later life as well, since later developmental landmarks also suggest slowing⁴. Such 'birth canal bottleneck' selection favouring slowed development is far removed from 'strong selection to form teeth quickly'. Where there is pleiotropy — overall developmental rates are a prime example of diverse body features controlled by a common set of genes — cause and effect are sometimes linked by long indirect paths that confound straightforward adaptationist reasoning.

WILLIAM H. CALVIN

University of Washington,
Biology Program, NJ-15
Seattle, Washington 98195, USA

1. Beynon, A.D. & Wood, B.A. *Nature* **326**, 493–496 (1987).
2. Isaac, G. Ll. in *Evolution from Molecules to Men* (ed. Bendall, D.S.) 528 (Cambridge University Press, 1983).
3. W.H. Calvin *The River That Flows Uphill: A Journey from the Big Bang to the Big Brain*. (Macmillan, New York, 1986).
4. Harvey, P.H. & Clutton-Brock, T.H. *Evolution* **39**, 559–581 (1985).

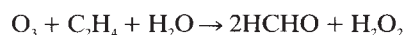
IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

The half-scale, radio-controlled flapping model of the pterosaur *Quetzalcoatlus* in flight over Death Valley. Wingspan of the model was 11m. Courtesy AeroVironment Inc.

Ozone and ethylene stress

SIR—In a recent News and Views article¹ "Adding ethylene to injury", M. Unsworth discussed the results of Mehlhorn and Wellburn², who reported that both endogenous ethylene (a stress-hormone) and exogenous ozone are co-operating prerequisites for injury to pea plants. This important experimental result is indeed a milestone in plant toxicology. But Unsworth's statement that a co-operation of ozone and ethylene in producing plant damage... "does not seem to have been suggested previously..." needs some comments.

In 1985, we reported³ changes in anti-oxidative activities in damaged spruce needles correlated with the ethylene precursor ACC or malonyl-ACC as a stress marker. We concluded that reaction products of the interaction between ozone and ethylene, namely peroxides and reactive aldehydes, might be the damaging species. We summarized these conclusions in the simplified equation:



The basis of this theory was published in 1984 in German⁴. In later publications^{5,6} we state that "the primary damaging reactions in spruce needles may operate as follows: (1) Trees under 'stress' produce the plant hormone ethylene; (2) ethylene and ozone react extremely fast, forming hydrogen peroxide and formaldehyde, compounds which may damage the wax layer; (3) ozone as a very aggressive..." (from ref.6). The findings of Mehlhorn and Wellburn with peas justify our earlier assumptions on forest decline and are therefore of utmost importance.

E.F. ELSTNER

*Institut für Botanik und Mikrobiologie
der Technischen Universität München,
Arcisstrasse 21, 8 München 2, FRG*

1. Unsworth, M. *Nature* **327**, 364–365 (1987).
2. Mehlhorn, H. & Wellburn, A.R. *Nature* **327**, 417–418 (1987).
3. Elstner, E.F., Osswald, W. & Youngman, R. J. *Experientia* **41**, 591–597 (1985).
4. Elstner, E.F. & Osswald, W. *Naturw. Rundschau* **37**(2), 52–61 (1984).
5. Osswald, W. & Elstner, E.F. *Ber. Deutsch. Bot. Ges.* **99**, 313–339 (1986).
6. Osswald, W. & Elstner, E.F. *Free Radical Res. Commun.* **3**, 185–192 (1987).

Sequence specificity of retroviral proteases

SIR—The protease encoded by retroviruses cleaves the *gag*-polyprotein to produce the amino terminus of the major core protein (p24, p27 or p30, depending on the virus)^{1–3}. On examining the amino-acid sequence that spans this cleavage site in available retroviral sequences we have found the strongly conserved pattern X-Y-Pro-Z, where X is generally small with some hydrophobic preference, Y is aromatic or large and hydrophobic, and Z

is small and hydrophobic (Fig. 1). The totally conserved proline forms the amino terminus of the major core protein. Some weaker clusters of conservation are also seen in the flanking residues. An additional closely related pattern of sequences is found in some *gag*-polyprotein sequences (Fig. 2), corresponding to the junctions of the p15 and p12 proteins, which may also be a site for cleavage by the viral protease. In the *pol*-polyprotein sequence of Rous sarcoma virus, a sequence matching this pattern occurs at the junction of the reverse transcriptase and the endonuclease, while in the AIDS viruses HIV-1 and HIV-2, a matching sequence occurs at the junction of the (presumed) carboxy terminus of the protease sequence itself with the amino-terminus of the reverse transcriptase, implying that the production of active reverse transcriptase in these viruses may well be dependant upon the action of the virally encoded protease. No matches with this sequence pattern are found at known cleavage sites in the various *env*-polyproteins. It is likely that this overall pattern represents the preferred amino-acid sequence for cleavage by retroviral proteases and may be of use in the design of specific inhibitors of retroviral protease activity for the chemotherapy of AIDS. Observations that retroviral proteases may be inhibited by thiol-specific reagents^{4,5} has led to suggestions that these

AKV-MuLV	Ser Ala Leu Tyr	Pro	Ala Leu	p15/p12 <i>gag</i>
Fr-SFFV	Ser Ala Leu Tyr	Pro	Ala Leu	
FeLV	Ser Ser Leu Tyr	Pro	Ala Leu	
Mo-MuLV	Ser Ser Leu Tyr	Pro	Ala Leu	
FeSV	Ser Ser Leu Tyr	Pro	Val Leu	
EqIAV	Ser Glu Glu Tyr	Pro	Ile Met	
SSV	Pro Pro Ile Tyr	Pro	Ala Thr	
RSV	Phe Glu Ala Tyr	Pro	Leu Arg	RT/enuc <i>pol</i>
HIV-1	Thr Leu Asn Phe	Pro	Ile Ser	Prt/RT <i>pol</i>
HIV-2	Ser Leu Asn Leu	Pro	Val Ala	

enzymes might be thiol-proteases⁶. A more related sequences have become available this hypothesis has appeared more unlikely since the position of a number of cysteine residues are found to be poorly conserved. In the recently published⁷ sequence of HIV-2, we have identified the likely position of the amino terminus of the protease as close to residue 85, with the carboxy terminus immediately preceding the probable amino terminus of the reverse transcriptase close to residue 185 (residue numbers are relative to the start of the *pol* open reading frame). No cysteine residues occur within this sequence.

MTV	Thr Phe Thr Phe	Pro	Val Val
HIAP-18	Gln Met Ala Phe	Pro	Val Phe
SRV-1	Lys Asp Ile Phe	Pro	Val Thr
HTLV-II	Thr Gln Cys Phe	Pro	Ile Leu
AKV-MuLV	Ser Arg Ala Phe	Pro	Leu Ala
Mo-MuLV	Ser Gln Ala Phe	Pro	Leu Ala
FeLV	Ser Gln Ala Phe	Pro	Leu Ala
H7BEV	Ser Ser Leu Phe	Pro	Leu Ala
HIV-1	Ser Gln Asn Tyr	Pro	Ile Val
HIV-2	Gly Gly Asn Tyr	Pro	Val Gly
ULV	Arg Glu Val Tyr	Pro	Ile Val
RSV	Val Val Ala Met	Pro	Val Val
FSV	Met Val Ala Met	Pro	Val Val
HTLV-I	Pro Ala Ile Leu	Pro	Ile Ile
FeSV	Ser Gln Ala Leu	Pro	Leu Ala
SSV	Thr Val Ile Leu	Pro	Leu Ala
BLV	Pro Ala Ile Leu	Pro	Ile Ile

Fig. 2 Alignment of other sequences in viral polyprotein which are cleaved to yield N termini and which may be substrates for the virally encoded proteases. Sequences are from AKV-MuLV, Friend spleen focus-forming virus (Fr-SFFV), FeLV, Mo-MuLV, FeSV, equine infectious anaemia virus (EqIAV), SSV, RSV, HIV-1 and HIV-2.

LAURENCE H. PEAR

*Cancer Research Campaign Biomolecular
Structure Unit,
Institute of Cancer Research,
Clifton Avenue,
Sutton, Surrey SM2 5PX, UK*

WILLIAM R. TAYLOR

*National Institute for Medical Research,
The Ridgeway,
Mill Hill,
London NW7 1AA, UK*

1. Yoshinaka, Y., Shames, R.B., Luftig, R.B., Smythers, G.W. & Oroszlan, S. *J. gen. Virol.* **66**, 379–383 (1985).
2. Yoshinaka, Y., Katoh, I., Copeland, T.D., Smythers, G. & Oroszlan, S. *J. Virol.* **57**, 826–832 (1986).
3. Kramer, R.A. *et al. Science* **231**, 1580–1584 (1986).
4. Yoshinaka, Y. & Luftig, R.B. *Cell* **12**, 709–719 (1977).
5. Dittmar, K. & Moelling, K. *J. Virol.* **28**, 106–118 (1987).
6. Yoshinaka, Y., Katoh, I., Copeland, T.D. & Oroszlan, S. *Proc. natn. Acad. Sci. U.S.A.* **82**, 1618–1622 (1985).
7. Guyader, M. *et al. Nature* **326**, 662–669 (1987).

Past influences

W.F. Bynum

The Development of American Physiology: Scientific Medicine in the Nineteenth Century. By W. Bruce Fye.

Johns Hopkins University Press: 1987. Pp. 308. \$38.50, £24.80.

Walter B. Cannon: The Life and Times of a Young Scientist. By Saul Benison, A. Clifford Barger and Elin L. Wolfe.

Harvard University Press: 1987. Pp. 520. \$30, £23.95.

THE historical role of experimental physiology in the development of modern medicine is nowhere better symbolized than in the name of the Nobel Prize for Physiology or Medicine. Many immunologists, pathologists, biochemists and molecular biologists have won the prize since it was first awarded in 1901, but its very name continues to reflect the extent to which scientific medicine in the late nineteenth and early twentieth century was perceived to rest on physiological foundations. In Britain and the United States (but less so in Germany and France), it was routinely the physiology department which first achieved independence from its traditional amalgamation with anatomy and then went on to spawn its own offspring such as biochemistry, bacteriology and experimental pathology.

Two recent books go some way towards providing an understanding of this phenomenon in the United States. Bruce Fye's monograph, timed to coincide with the centenary of the American Physiological Society, surveys the fortunes of experimental physiology and physiologists from about 1860 until the century's end, by which time Walter B. Cannon had begun his own lifelong association with the Department of Physiology at Harvard. Because Cannon was such a key figure at Harvard, his biography offers a detailed case study of one institution where the issues surveyed by Fye were the subject of much debate and sometimes acrimonious division. The Harvard example reinforces the suspicion that Fye's account, useful though it is, is too dependent on rhetoric and intention. The reality and practice were more complicated.

More than half of Fye's book is given over to biographical studies of four individuals: J.C. Dalton at the College of Physicians and Surgeons in New York, Silas Weir Mitchell in Philadelphia, Henry Bowditch at Harvard and Newell Martin at Johns Hopkins. Dalton is hardly remembered today, Mitchell is better known for his literary and clinical work than his physiological research, and neither Bowditch nor Martin was a first-class experimentalist. Mitchell in fact was unsuccessful in his attempts to obtain a chair in physiology at the University of Pennsylvania and Jefferson Medical College, although his researches on the

effects of various snake venoms on animals were as well executed as anything achieved by any of the handful of his contemporary professional physiologists in the United States.

It may be that Fye is aware that none of these first-generation experimental physiologists was of world standard, and that their lasting significance lay in their commitment to what he calls the 'research ethic'. But I missed in this account an analysis of how their espousal of a research ethic manifested itself, not simply in their



Walter Cannon — new age in physiology.

own experimental work, but in their actual teaching, calls for curriculum reform and relationships with clinical colleagues: themes which are richly explored by Benison, Barger and Wolfe in the instance of Cannon and Harvard.

When Cannon arrived from Minneapolis in 1892 as an undergraduate at Harvard, Bowditch was already a venerable figure whose time was divided between the physiology department (where he taught primarily by lecture and demonstration rather than hands-on practical work) and the wider world of Harvard medical politics. The practical teaching had been largely turned over to William Porter, a man whom Cannon was

later to find difficult but who united with Bowditch in encouraging the young man's experimental bent when, following his graduation from Harvard College, he stayed on to study medicine at the medical school. The decision not to enter medical practice was not easy, but already as a medical student he had developed a talent for experimentation. Indeed, his classic studies on mechanical aspects of digestion, using the newly-discovered X-rays and contrast substances such as bismuth subnitrate, were begun in the first year of his medical course. Unsurprisingly, Cannon was asked to join the physiology department after he took his medical degree, retiring from it almost half a century later.

Saul Benison is a historian, Clifford Barger a physiologist and Elin Wolfe an archivist. Together they have produced a biography of the first 45 years of Cannon's life which is a model of its kind, reflecting a unity of style and purpose. It is historically sophisticated, particularly in its interweaving of issues facing American higher education with Cannon's own participation in debates about science, medicine and research. It is equally good in its treatment of Cannon's own research, sympathetically analysing his work, especially in digestion and on the functions of the adrenal medulla, within the context of his contemporary scientific community. Ultimately, its strengths lie squarely on its archival grounding. This biography makes use of no fewer than 54 manuscript collections; centrally, of course, the Cannon papers at Harvard but much else as well. Unpublished correspondence, diaries, reports and other documents give it an immediacy and authority which is refreshing.

Three themes link the Cannon biography with Bruce Fye's survey of earlier developments. The first is vivisection in practice and antivivisection in sentiment. The use of experimental animals was something which set the new-style physiologists off from their predecessors (who in their own times had still claimed to be 'scientific'). Vivisection brought in its wake a reaction, perhaps strongest and most organized in Britain, but vigorous and visible in the United States as well. The 1876 Act which regulated vivisectional research in Britain was national; in the United States it was generally the State Legislatures that antivivisection groups sought to influence. This made for a more complicated scene, though, as in Britain, experimentalists banded together nationally in an attempt to proffer a common front. Cannon spent many hours defending experimental physiology, and antivivisectionism is a leitmotiv in Fye's monograph as well. The two volumes do much to clarify the American situation before the First World War.

Freedom from antivivisection pressures

was one precondition for experimental medicine. Endowments were another, and the importance of philanthropy at Harvard, Johns Hopkins and elsewhere is a second central theme in both books. Johns Hopkins has been left well endowed by its founder, but even there its president and research-minded faculty sought outside funds to build new departments and laboratories, buy new equipment and increase staffing levels so that research time and facilities would materialize. The support of medical research by the Rockefeller and Carnegie Foundations is well known, but smaller foundations and individuals were also important, and these two studies reveal the extent to which early experimentalists were forced to cultivate potential benefactors.

A third issue revolved around what was called the 'full-time system'. Most experimental physiologists agreed that their discipline needed men whose time was not diluted by medical practice, and Fye suggests that by the turn of the century the battle for full-time medical scientists was essentially won. A final judgement on that would require looking beyond the dozen-or-so elite medical schools on which he

focuses, and the example of Harvard suggests the extension of the full-time concept to academic clinicians was more fraught than his optimistic conclusion implies. The history of career structures within medicine and medical science is a subject ripe for systematic exploration.

These two volumes can be fruitfully read together, and we must hope that Cannon's later career will be dealt with by the team which has given us the portrait of the young scientist. When he left for France in May 1917, there was still much scientific, professional and political work before him, including *The Wisdom of the Body* (1936), that splendid popularization of the concept of homeostasis, a word he coined. In the mid-1960s, the physiology department at Yale urged all its medical students to read Cannon's book. I have always been glad I took the advice and look forward to learning more of the circumstances of its composition, and the research on which it was based, from the trio who have so beautifully brought his early career to life. □

W. F. Bynum is at the Wellcome Institute for the History of Medicine, 183 Euston Road, London NW1 2BP, UK.

A bit on one side

David Caplan

Cerebral Lateralization: Biological Mechanisms, Associations and Pathology. By Norman Geschwind and Albert M. Galaburda. MIT Press: 1987. Pp.283. £26.95, \$30.

THE title of this book is somewhat misleading. Although the focus of Geschwind and Galaburda's work is on cerebral lateralization, the text is much wider in scope. It deals with the structural asymmetries in nervous tissue, which the authors argue underlie functional lateralization, and goes on to develop a theory of the genetic and intrauterine environmental determinants of the structural asymmetries. This leads the authors to consider asymmetries as far back in the phylogenetic scale as single-celled species and to invoke basic asymmetries in elementary subatomic particles (for instance, the direction of spin in neutrinos) as relevant to the determinants of structural asymmetry and functional lateralization. In addition, the authors report an enormous number of associations of what they call "anomalous dominance" — the situation in which an individual is not strongly right-handed. These associations include numerous forms of learning disabilities, but also, according to the authors, a wide variety of somatic abnormalities, especially diseases of the immune system.

The reader will find in these pages

extensive documentation of the extent of asymmetries throughout the phylogenetic and ontogenetic scales, both in the central nervous system and elsewhere. A major strength of the book is the richness of the documentation of these features of neurobiology. Almost all readers will find something about functional and anatomical asymmetry that they did not previously know. For instance: in most birds and some bats only one ovary is functional (usually the right); in several species, the right hypothalamus has twice the LHRH (luteinizing hormone-releasing hormone) content of the left; the siamese cat (an albino) has an anomalous pattern decussation of the optic nerves; the rate of turning to the left in response to amphetamine is higher in female rats from litters with higher percentages of male siblings. This book presents a truly impressive collection of facts about functional lateralization, structural asymmetries, and their associations. Geschwind's enthusiastic curiosity and Galaburda's attention to detail combine to make the book fascinating reading.

Where I found this volume to be considerably less satisfying, however, was in the theory that the authors develop to tie together the facts they report. As far as I can see, Geschwind and Galaburda have not enunciated a single theory with a deductive structure, but rather a host of potentially unrelated postulates regarding lateralization, its determinants, and its associations and their determinants.

What I take to be the essence of this theory, as it pertains to human functional

lateralization, is the following. Genetic factors are responsible for determining the pattern of handedness in humans coding for features of the left hemisphere that make for right-hand preference in the majority of individuals and for random dominance in a significant minority of individuals. Intrauterine hormonal effects greatly alter the neural structures which determine handedness. The hormone of greatest influence is testosterone, which slows down the development of cerebral cortex. Testosterone exerts its effect primarily on the left hemisphere because maturation of the right hemisphere is more complete at the point at which intrauterine testosterone levels rise. Testosterone levels (both intrauterine and postnatal) also have major effects upon other systems, in particular the immune system where they also retard development and accelerate degeneration. A very large number of ancillary hypotheses and predictions are related to and sometimes follow from these basic assumptions.

One of the reasons that this speculative theoretical aspect of the book is not satisfying is that crucial data supporting these key assertions are missing. For instance there is no evidence supporting the assertion that testosterone preferentially retards the growth of cortical areas of the left hemisphere. Although it is true that there are more left-handed men than women (which is consistent with this hypothesis), there is no evidence that there are 'handedness' areas of the brain that they are larger in the left hemisphere or that they are proportionally smaller in men than in women. What is known is that the language areas of the brain are asymmetrical (we are greatly indebted to Geschwind and Galaburda for the documentation of these asymmetries), and that the size of these asymmetries varies as a function of handedness. There is no evidence, however, that the size of these asymmetries varies as a function of sex. Indeed, to the extent that available evidence is reliable, it indicates that sex differences in the functional lateralization for language (and, presumably, the corresponding cortical asymmetries) go the opposite way to that predicted by Geschwind and Galaburda's theory (McGlone, *J. Behavioral and Brain Sciences* 3, 215–263). At many points in the book, the authors support one speculation by reference to another, inadequately supported, assumption. What is frustrating about this volume is that, despite the enormous numbers of facts reported, many theoretical statements lack adequate empirical support.

I do not wish to suggest that problems such as these negate the entirety of the theoretical work presented in the book. Indeed, no single empirical problem could negate the entirety of the authors' formulation, as it consists of a great number

of assertions, assumptions, and predictions, many of which are quite independent of others. Many ingenious and possibly brilliant speculations are to be found in this set. For instance, the authors view fusion of the two hemispheres as occurring along a continuum whose (functionally deleterious) endpoints are complete fusion (holoprosencephaly) and complete separation (agenesis of the corpus callosum) and in which many degrees of 'balance' of 'lateral dissociation' occur in normal individuals giving rise to variation in functional capacities. This is a novel and interesting idea.

This volume thus presents an extensive array of facts regarding functional lateralization, asymmetry and associated con-

ditions. A highly speculative theoretical formulation accompanies these data, but the scientific value of this formulation is hard to evaluate because of the multiplicity of postulates in the theory, and their tangential relationship to the many facts presented. Despite this, the volume presents enough data to suggest strongly that many of the phenomena described are related, and it emphasizes the need for understanding of the cellular, subcellular, and physical bases for asymmetry and lateralized function. It should stimulate research into these topics. □

David Caplan is Associate Professor of Neurology and Adjunct Professor of Linguistics, Montreal Neurological Institute, 3801 University, Montréal, Québec, Canada H3A 2B4.

Spotting gels

S. P. Spragg

Two-Dimensional Electrophoresis and Immunological Techniques. By Bonnie S. Dunbar. Plenum: 1987. Pp.371. \$59.50.

It is eleven years since O'Farrell published his widely acclaimed paper describing the procedure for two-dimensional electrophoresis of proteins, and about 60 years since Tiselius first separated mixtures of proteins by electrophoresis. I wonder what Tiselius and his contemporaries (such as Longworth) would feel about the present techniques and their applications? Possibly they would miss interpreting the results in terms of electrical constants for the molecules. Instead of measuring boundaries and converting the movements to mobilities, the experimenter now ends up with visually pleasing two-dimensional patterns of spots, each being a separate peptide or protein. Unfortunately, the end-products of electrophoresis are often not as 'clean' as shown in the literature, and a book such as this laboratory manual for novices is to be welcomed.

The contents are divided into two distinct sections. The first, an introductory stroll through results and previously reported procedures, is not cluttered with details of reagents or apparatus: these are left for the second section. The author covers many of the applications and gives references which the beginner will find useful. These references are not intended to be complete, but as reviews are included this is not a serious deficiency. The coverage of the physicochemical basis of electrophoresis is brief, almost to the point of being naive, but possibly this is all that is required for the experiments described. The very useful chapter on trouble-shooting will need to be revised regularly because some of the artefacts known today are not mentioned. It is a pity more information was not given on the analysis of the patterns — gels are fragile and difficult to store, and yet are the primary records, so leaving photographs and processed computer images as the only permanent experimental records.

The second section contains detailed recipes for the preparation of reagents, treatment of proteins to identify components, and photography. Many important experimental details are mentioned making it the lasting part of the book. Most biochemical laboratories will be able to follow the recipes without much trouble.

I feel this book does indeed fill a gap. Laboratories using two-dimensional electrophoresis to separate complex mixtures of proteins would do well to acquire it. □

S. P. Spragg is a Senior Lecturer in the Department of Chemistry, University of Birmingham, PO Box 363, Birmingham B15 2TT, UK.

Scattered matter

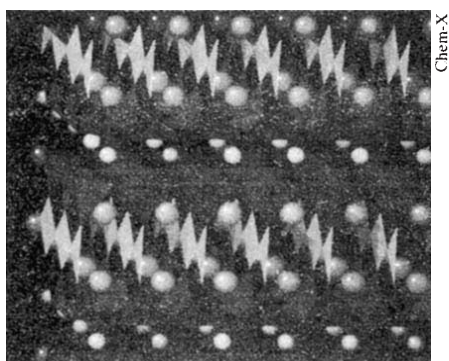
Stephen W. Lovesey

Fifty Years of Neutron Diffraction: The Advent of Neutron Scattering. Edited by G.E. Bacon. Adam Hilger: 1987. Pp. 280. £30, \$66.

THE Commission on Neutron Diffraction of the International Union of Crystallography (IUC) is fortunate to have enlisted the distinguished experimentalist G.E. Bacon to edit the volume *Fifty Years of Neutron Diffraction*, which it conceived and encouraged. Two previous volumes for the IUC commemorate the fiftieth anniversaries of X-ray (1962) and electron (1981) diffraction. The scattering of radiations (taken to include both elastic diffraction and spectroscopy) are techniques of great importance in modern condensed matter and materials science research, whose results pervade almost all aspects of condensed matter physics, chemistry and biology. Moreover, the techniques continue to improve in quality and to expand in their range of applications. For example, the recent development of intense beams from synchrotron light sources opens the vista of magnetic photon diffraction, whereas previously magnetic structure investigations were conducted only with neutron beams.

Neutron scattering is an intensity-limited technique. In consequence, improvements in spectrometer design and source intensities and characteristics have a very marked effect on the impact of neutron scattering. Today researchers are beginning to reap benefits from spallation neutron sources, which are based on proton synchrotrons, and the development of the next generation of reactor sources is in hand.

Although these current projects are accorded some attention in the volume, most of the 42 contributors write about earlier developments. Some have as their



Current progress — crystal structure of a superconducting ceramic determined by high-resolution neutron powder diffraction.

subject the growth over the years of local user communities. On the other hand, B.N. Brockhouse and F. Mezei, for example, describe the research and development for the spectrometers they pioneered. The book includes reminiscences by some of the founding fathers of the technique, discussions of world-wide growth, and mini-reviews of some major areas of application and the impact of so-called high flux reactors. The material is unlikely to appeal to readers without a vested interest in the subject.

A noticeable omission is a survey of the essential part played by theoretical physicists in the development of the subject; for example the two papers by L. Van Hove, published in 1954, can justifiably be argued to be the cornerstone of the interpretation of experimental data. Also absent amongst the contributors are such great scientists as R.D. Lowde, R. Nathans and G. Shirane. However, one suspects that the likes of Shirane are too hard pressed with the demands of current research for the more leisurely task of mulling over the past. □

Stephen W. Lovesey is a senior staff researcher at the Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, UK, and visiting professor of physics at Southampton University. He is the author of the two-volume work Theory of Neutron Scattering by Condensed Matter (Oxford University Press, 1986).

Star quality

Gary Steigman

Highlights of Modern Astrophysics: Concepts and Controversies. Edited by Stuart L. Shapiro and Saul A. Teukolsky. Wiley: 1986. Pp.389. £34.40, \$37.50.

In October 1984 some 400 scientists gathered on the campus of Cornell University to honour Edwin E. Salpeter on the occasion of his sixtieth birthday. Without doubt, Salpeter has been in the forefront of many of the most important developments in astrophysics for more than 30 years. That "he is responsible for much of the progress achieved in theoretical astrophysics in the past 30 years", may be hyperbole. Then again, perhaps not. As the 13 articles in this collection continually remind one, Salpeter's influence on and contributions to modern astrophysics have been pervasive.

The articles in this collection range in subject matter from nuclear to neutrino astrophysics, from black holes to the Universe, from observational astronomy to philosophy. The authors are all acknowledged leaders in their fields. Inevitably, some of the contributions have appeared elsewhere, and little new is to be gleaned from the version that appears in this volume.

This, however, is a quibble. In general, the quality of the writing and the depth of detail presented recommends this book very strongly. But I must part company with the editors who think that "non-professional readers with little technical background . . ." should read it; the non-expert will quickly be left in the dust. However, professional astronomers or astrophysicists will find this volume an excellent resource for acquainting themselves with research in allied fields; for the astronomy/physics student, many of the articles will provide comprehensive and comprehensible introductions to some of the most exciting topics in modern astrophysics. Indeed, because of the pedagogical emphasis of many of the articles, the book has not suffered from the two years delay in seeing the light of day. Nor, for this reason, will its usefulness decrease very quickly, even in our ever-changing rapidly developing fields of modern astrophysics. In the wake of Supernova 1987a there is an eery timeliness in Bethe's monograph on supernova theory and Ruderman's on neutron stars. Are you interested in knowing what was predicted and what observers might expect to find in the coming months? Read those articles. Has something been shot from the supernova at relativistic speeds? Read Martin Rees's article on jets and Marshall Cohen's on relativistic motion. Can we do Cosmology with the supernova? Read

Wagoner's contribution. If irony appeals to you, turn to Riccardo Giacconi's overview of the yet to be launched Space Telescope and contrast it with Freeman Dyson's persuasive "Quick Is Beautiful".

Whenever the writings of 13 outstanding scientists are assembled, variety of style and content will be the rule. These pieces are no exception. Those who find historical perspective appealing should try Willy Fowler's detective story of carbon and oxygen synthesis (and note Hoyle and Salpeter's complementary contributions), or George Field's historical overview of interstellar physics. Do you seek a pocket course in cosmology or in black hole Physics? Proceed directly to the comprehensive and detailed articles by Wagoner or Kip Thorne. Where do we stand on the Solar Neutrino Problem or the Dark Matter Problem? Bahcall and Rubin will tell you.

Although there are the inevitable typographical errors, the quality of production is, with one exception, excellent. The

reproduction of photographs and multi-colour maps is useless; the original glossies should have been included, or the figures revised or eliminated.

This is a valuable book because it can be used in many ways. Some articles provide tutorials, some overviews. All have extensive bibliographies, and there is a good, overall index. But, it is not relaxing summertime reading; with the exception of the "Perspectives" articles by Philip Morrison and Freeman Dyson, to profit from this book requires work. But that work will be rewarded. This volume is a fitting tribute to the accomplishments of Ed Salpeter; to his contributions and influence on the directions modern astrophysics has taken and its successes. □

Gary Steigman is Professor of Physics and Astronomy at The Ohio State University where he is building a group of theoretical astrophysicists to work at the frontier of cosmology and high energy physics. Steigman has worked on problems of the physics of the interstellar medium, on Big Bang nucleosynthesis and on the particle physics/astrophysics connection.

A big fat book

A. G. Lee

The Physical Chemistry of Lipids: From Alkanes to Phospholipids. Handbook of Lipid Research, Vol. 4. By Donald M. Small *et al.* Plenum: 1986. Pp.472. \$89.50.

LIPIDS are an unusual class of biological molecule in that they are insoluble in water. They are also an unusually heterogeneous group of compounds, including hydrocarbons, steroids, long-chain alcohols and acids, glycerides, phospholipids and glycolipids. Their biological interest lies in the uses to which their insolubility can be put by the cell, principally, of course, in the formation of biological membranes but also in the formation of lipoproteins necessary for transport about the body. Because they exist in a solid, undissolved state in biological systems, the properties of the solid state are of considerable importance, and this is the principal concern of this book.

For a biologist, the most interesting lipids are probably the phospholipids, glycolipids and sterols which make up the biological membrane. The properties of the phospholipids and glycolipids are best understood in terms of the properties of simpler, long-chain alkyl derivatives such as the alkanes, alkyl alcohols and fatty acids. Information about these compounds is widely scattered throughout the scientific literature and a major virtue of Small's book is that he has sorted through much of this rather tedious stuff, rendering it into manageable form. Unusually, the book is largely by the one author (with contributions from Craven, Lange,

Shipley and Steiner), which makes for an integrated and uniform approach.

The various modes of packing of hydrocarbon chains are particularly well described, as are the wide variety of different solid phases which these compounds can adopt. A clear discussion of the conditions under which these various phases are formed is valuable because of the current interest in the non-bilayer phases formed by phospholipids and their possible importance in the functioning of biological membranes: there is no doubt that a wide variety of non-bilayer phases can be formed under the appropriate conditions, but it is not so obvious that these phases are exploited in any way in biological systems; on the contrary, it seems that they are avoided at all costs.

Small's enthusiasm seems to have dwindled somewhat on approaching the vast literature on phospholipids, and the level of coverage here is less comprehensive than that given to the simpler lipid molecules. Furthermore, in places the literature coverage extends only to the early 1980s, suggesting a long gestation period for the book. But here begins and ends my nitpicking.

This is undoubtedly a true handbook which will be the first source to turn to for data about the physical properties of lipids, particularly the simpler lipids. It is a mine of information. Where else could one find out the solubility of alkanes in water or water in alkanes, the melting point of tsuzuic acid or the structure of obtusilic acid? Small deserves all our thanks for making so much basic data available in so clear and manageable a form. □

A.G. Lee is a Lecturer in the Department of Biochemistry, University of Southampton, Southampton SO9 3TU, UK.

A weapon for the twenty-first century

D. M. Ritson

One component of the US Strategic Defense Initiative is the nuclear-explosion-pumped X-ray laser. Although there are many obstacles to its development, such a weapon is technically feasible.

IN his 'Star Wars' speech of March 1983, President Reagan announced plans to build a defensive system to protect the United States against attack by intercontinental ballistic missiles. The Strategic Defense Initiative (SDI) envisages a multilayered defence system, with the first line of defence destroying missiles just after launch while still in their boost phase. A candidate for this first line of defence is the nuclear-explosion-pumped X-ray laser.

The events leading up to this speech are now well known. In 1982, at the thirtieth anniversary of the Lawrence Livermore Laboratory, Edward Teller, in a speech seven months before the President's announcement, said: "What this lab can accomplish now is more important than what we have ever accomplished before. The third-generation effort [on nuclear technologies] gives us every expectation of an effective nuclear defense, and if defense by nuclear weapons is possible, we must have it". The main item referred to was the development of a nuclear-pumped X-ray laser with a projected capability of multiple destruction of missiles in the first few minutes after launch at ranges in excess of 2,000 km. Teller followed up this speech by letters to the President, and in a personal visit he apparently persuaded President Reagan to proceed with SDI.

Whatever the nuclear-pumped X-ray laser's potentialities may or may not be for defensive purposes, should it achieve even a fraction of its projected performance it would certainly provide a basis for offensive anti-satellite weaponry. Neither the United States nor the Soviet Union can sit by and watch its development by the other side.

Although details of nuclear-pumped X-ray lasers have been known to both the US and the Soviet military establishments for over seven years, even basic information remains classified. Publicly, the main information until recently was provided by articles in *Aviation Week and Space Technology*, including the 1981 'announcement' of the first demonstration of nuclear-pumped lasing¹. (Any lingering doubts as to the success of that demonstration have been removed by the recent report of the American Physical Society on directed energy weapons²).

Despite classification, enough information is available to outline the objectives,

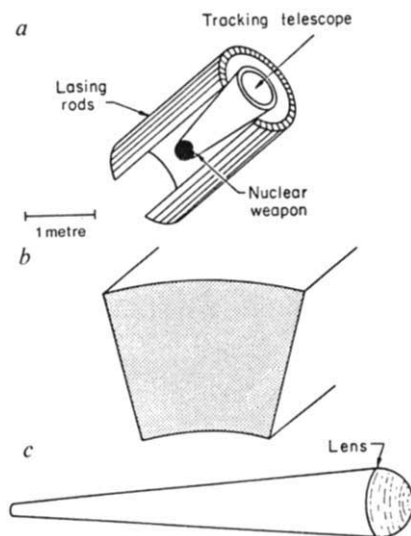


Fig. 1. *a*, Schematic layout of the nuclear-pumped X-ray laser. The fission device is surrounded at a radius of one metre by a shell of laser 'modules'. *b*, End view of one of the modules, which is built up from individual zinc lasing wires embedded in a plastic matrix. *c*, A highly simplified view of a lasing wire designed to overcome geometrical optics limitations with the use of a 'lens'.

design principles and progress of the nuclear-pumped X-ray laser project in the United States. The first section of this article describes the parameters and layout of a conceptual laser device. Discussion of the underlying physics that determines the choice of parameters and design follows, with a brief summary to conclude.

The reader is cautioned that many design problems are unresolved, and present performance is as yet orders of magnitude from that required for a militarily useful weapon¹. Further, I do not examine tactical questions here, or the defensive or offensive roles that laser weaponry could assume in a first-strike attack. (For discussion of these matters see ref. 3.)

The conceptual layout. The objective of the nuclear-pumped X-ray laser, when used in weapon systems, is to direct much of the energy released in a nuclear explosion into a very narrowly collimated beam or beams. The beam(s) could then be focused upon and destroy or damage targets at very long ranges. X-ray laser beams are strongly attenuated by matter, thus effectively limiting their use to outside the atmosphere. In addition, the initial nuclear explosion would destroy the

device, thus precluding subsequent use.

Figure 1a shows the overall layout. At the centre is a 120-ktonne nuclear device to serve as the pump source of X-rays. The fission device is in turn surrounded by 'laser modules', each 2 metres long and set around a cylinder 1 metre in radius (for discussion purposes there are assumed to be 50 such modules). The modules are made up of a large number of parallel (zinc) wires, 0.1 mm in diameter at the output end, embedded in a plastic matrix. Each individual zinc wire is a lasing element. Figure 1b shows an end view of a module. Figure 1c shows a simplistic design of a laser wire, in the form of a cone with a 'lens' at the end for focusing. (In practice focusing is likely to be achieved by the use of non-uniform materials to create refractive index gradients in the ionized plasma.)

The pump source first ionizes the atoms in the wires, forming an ionized plasma transparent to X-rays. The hot, high-pressure plasma rapidly expands to equalize the plasma pressures in the wires and plastic matrix, cools and then lases, with the pump source providing and maintaining the required level-inversion density. Appropriate shaping of the wires is assumed to provide sufficient changes of refractive index to provide a focused output beam from individual laser elements. Groups of laser elements can be directed at different targets, if desired, thus permitting several targets to be attacked simultaneously.

Table 1 lists the relevant parameters. The choice of parameters and the basis of the estimates of potential performance are discussed in the following sections.

Useful units and destructive power. The energy release corresponding to a 120-ktonne bomb is 4.8×10^{14} J. If the pump-power conversion efficiency to laser energy were 0.1% and laser beams could be directed into a solid angle of 10^{-10} steradians (the value used in Table 1), then the resulting energy flux F onto each of four targets would be $F = [(1 \times 10^{-3})(4.8 \times 10^{14})]/(4 \times 10^{-10}) \approx 1.2 \times 10^{21}$ J sr⁻¹, which is a level estimated as sufficient to destroy a booster at 2,000 km⁴.

This output yield is often expressed in terms of an enhancement factor, G , the ratio of the energy fluxes per steradian of the directed laser beam to that produced in the initial explosion, where the energy spreads isotropically over 4π steradians.

With the above numbers, $G = [(4\pi)(1 \times 10^{-3})]/(1 \times 10^{-10}) = 1.2 \times 10^8$.

Because the flux of energy from a source decreases inversely as the square of the distance from the source, this enhancement factor corresponds to the full collimated laser beam being equal in destructive power to the explosion of the driving nuclear device at a distance $G^{1/2}$ or $\sim 10^4$ times closer to the target.

Characteristics of nuclear fission devices⁵. A nuclear explosion is driven by a chain reaction of neutrons causing and arising from fission. The explosion requires about 50 generations. Over half of the energy released will be associated with the last few generations of the chain reaction and will occur in a time of $\sim 10^{-8}$ s. About 70% of the energy release is in the form of prompt X-rays with an approximately thermal equilibrium black-body spectrum and a mean photon energy of ~ 10 keV.

Absorption of high-intensity X-ray beams. X-ray beams of low intensity traversing solid material are strongly attenuated because of photoelectric absorption, with an exponential fall-off of intensity with path length. The attenuation coefficients vary strongly with frequency and the atomic number (Z) of the material.

The attenuation of a very intense beam of X-rays traversing matter is different. A fraction of the energy of the beam will be subtracted, enough to strip the atoms of their bound electrons. After stripping, in the absence of strong recombination processes, the material will cease to absorb appreciably. This process is often referred to as 'bleaching'. A fixed amount of energy per unit volume is required to bleach the material. From this follows the basic equation for the attenuation of a high-intensity X-ray beam traversing material: $dI/dx \approx K$, where I is the beam intensity (in J cm^{-2}) and x is the thickness of material traversed. K is independent of density variations in the material if thicknesses are expressed in terms of density $\times$ length; accordingly, x is in units of g cm^{-2} . K will be a constant that takes into account the atomic properties of the material and is relatively independent of the spectral composition of the beam.

The thickness R of material needed to absorb the bulk of the energy in the beam is $R = I/K$.

The numerical value for K can be estimated as follows. For low-density materials such as air, the approximation that only a small fraction of atoms suffer recombination within the pulse time of 10^{-8} s is valid. The energy required for bleaching is then the energy required to ionize the atoms and to supply kinetic energy to the plasma. Equilibrium is established both by the photoionization processes and through subsequent processes caused by the primary photoelectrons,

Auger-emitted electrons and secondary spectroscopic transitions. Provided that the primary X-ray energies are large enough to ionize even the most tightly bound electrons, the final equilibrium state in which the material is left is almost entirely dependent on the total energy deposited in the material and independent of the X-ray frequency. The energy required to ionize an atom of nitrogen or oxygen in air is the binding energy of ~ 1 keV (see, for instance Table 8-2 in ref. 6). With the assumption that an additional 50% of energy is required to provide kinetic energy to the plasma, ~ 1.5 keV is required to bleach an atom of oxygen or nitrogen in air. The value of K for air is then $\sim 1 \times 10^7 \text{ J g}^{-1}$.

Some numerical examples of typical penetrations through the atmosphere to be expected for a weapon with the output energy to destroy a missile at 2,000 km ($10^{21} \text{ J sr}^{-1}$ at 2,000 km) are of interest. At 2,000 km the beam will be effectively absorbed after penetrating $\sim 2.5 \times 10^{-3} \text{ g cm}^{-2}$ through the atmosphere. This corresponds, for a beam directed vertically downwards from a laser in orbit 2,000 km from the Earth's surface, to absorption after penetration to a height of 120 km above the Earth's surface.

Although the widely diffused beam at the end of range will penetrate only a small layer of the atmosphere, the tightly collimated beam close to the firing point can penetrate substantial thicknesses. For instance the beam from an X-ray laser directed vertically upwards, from a height 55 km above the surface of the Earth, can penetrate through the remaining 1.7 g cm^{-2} of atmosphere with only a 25% loss of intensity. This leads to the

remarkable property that ground-launched X-ray lasers can attack opposing space-based X-ray lasers without fear of retaliation. (The conclusions of this section are insensitive to variations in wavelength. For X-ray wavelengths longer than $20 \text{ \AA}$, corresponding to the K -absorption edge for nitrogen atoms, the bleaching constant K is decreased to $\sim 1.5 \times 10^6 \text{ J g}^{-1}$. This, however, only increases penetration depths by ~ 15 km, because of the exponential dependence of atmospheric density on height.)

Mechanical effects caused by the absorption of the pump flux. The absorption of pump-flux energy E_0 transfers an outwardly directed momentum $p = E_0/c$ (where c is speed of light) into the laser assembly. The resultant velocity, equal to p/m (where m is the total mass of the assembly) is, however, small and will not be considered further.

The important effects are caused by the pump flux fully ionizing the laser assembly and almost instantaneously converting it into hot, ionized plasma. For the zinc wires the initial temperature of the plasma will rise to $\sim 10^7 \text{ K}$ and the pressure to $\sim 10^9 \text{ atm}$. Both the lasing wires and the plastic in which they are embedded will be converted to hot, ionized plasma. The initial density of electrons in zinc is seven times larger than in plastic, and thus the plasma pressure of the ionized zinc will be correspondingly larger than the plasma pressure in the plastic. The zinc plasma will expand and the 'plastic' plasma will be compressed until the gas pressures equalize. The final equilibrium configuration will correspond to nearly equal electron densities in the two plasmas. The assembly under consideration would start

Table 1 Conceptual design parameters for a nuclear-pumped X-ray laser

Parameter	Value
Fission energy (kilotonne)*	120
Laser material†	Zn
Atomic number of material	30
Spectroscopic transitions	$5f \rightarrow 3d$
Laser wavelength ($\text{\AA}$)	14.2
Pump energy conversion efficiency	$\sim 0.1\%$
Solid angle containing the laser beam (steradian)	$\sim 10^{-10}$
Radius of laser assembly (m)	1
Length of wires (m)‡	2
Diameter of lasing wires (mm)§	0.2
Number of wires	$\sim 10^4$ – 10^5
Weight (kg)	~ 200
Number of independent targets	4
Total output intensity at 2,000 km range (J sr^{-1})	$\sim 4 \times 10^{21}$

* The power level chosen is arbitrary and could be increased by large factors. For convenience we talk about a fission device. There is no obvious reason why fission-fusion sources would not function equally well at higher power levels.

† Zinc is chosen for convenience. Elements with atomic numbers from 20 to 40 are all possible choices.

‡ Long laser elements are required both to provide sufficient gain for lasing and to achieve good output collimation.

§ The diameter given is the expanded diameter and it is assumed that focusing is incorporated. In the absence of focusing the wires would have an expanded diameter of $5 \times 10^{-3} \text{ cm}$, and a collimated intensity lower by a factor of ~ 25 .

|| This is believed adequate to target four boosters⁴.

with 92% of the volume in the form of 'plastic' with specific gravity 1 and 8% in the form of zinc wires with specific gravity 7.1. The new expanded state will have specific gravity of 1.5 for both the zinc and 'plastic' plasma, corresponding to a fivefold expansion for the zinc.

The local expansion rate of plasma proceeds with a velocity v close to that of an acoustic shock wave: $v \approx \sqrt{2E/M}$, where E is the average kinetic energy per atom (~ 30 keV) and M is the atomic mass of $\sim 60 m_p$ (where m_p is the proton mass) giving $v = 3 \times 10^7$ cm s $^{-1}$. This corresponds to expanding to local equilibrium densities in $\sim 10^{-9}$ s, a time short compared with the $\sim 10^{-8}$ -s duration of the pump pulse.

The unbalanced pressures generated by the conversion to plasma will also cause expansions and distortions of the overall assembly. A pressure difference ΔP acting on a thin shell of material, with mass M per unit area, causes an acceleration $\ddot{x} = \Delta P/M$, and a displacement in a time Δt of $\Delta x = \ddot{x}(\Delta t)^2/2$.

Integrated over the 10^{-8} -s duration of the pump pulse, this leads to an estimated expansion of $\sim 10^{-3}$ cm for a 10-cm-thick laser assembly. If this expansion occurred uniformly over the volume, the angular spread of the laser beam would remain unaffected. However the energy deposition will be inhomogeneous in both space and time over the 2-m length of the laser wires and non-uniform displacements comparable to this $\sim 10^{-3}$ cm might be expected. This would cause a degradation in angular definition of the laser beam by $\sim 10^{-3}/200$, or 5×10^{-6} radians within design tolerances. A further source of degradation in collimation could occur if the inhomogeneous energy absorption generated coherent acoustic waves within the plasma. Consideration of such complex and detailed plasma physics is outside the scope of this article.

The X-ray lasing process (see also refs 7, 8). Before lasing can occur, the material must be fully ionized by the pump flux so that X-ray absorption in the material is small. Attainment of lasing requires the creation of a non-equilibrium population inversion. The presence of an inverted population leads by stimulated emission to amplification of a photon flux, and if the gain is sufficient, to lasing. For X-rays specular reflection can be obtained only at grazing incidence, and therefore, unlike conventional lasers operating in the visible region, lasing must be achieved in a single pass of the photons through the lasing element. For lasing to occur in a single pass, the amplification factor (small signal gain per unit length multiplied by the length of the lasing element) must exceed ~ 25 . The amplification factor is directly proportional to the inversion population density, inversely proportional to the line width and directly proportional to the length of the lasing element.

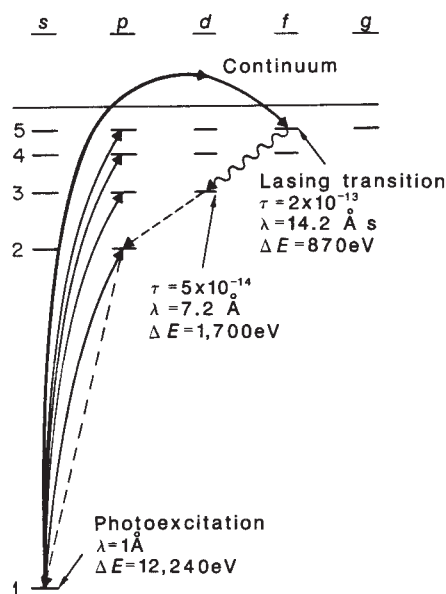


Fig. 2. Spectroscopic level diagram for 'hydrogenic' zinc. The projected lasing level¹⁰ ($5f \rightarrow 3d$) is indicated by the wavy line. The solid lines show the reabsorption and photoionization transitions from the $1s$ level. The expected lifetime for the $5f \rightarrow 3d$ transition is 2×10^{-13} s and for the $3d \rightarrow 2p$ transition 5×10^{-14} s thus making possible a population inversion.

In more detail, the process of nuclear-pumped X-ray lasing can be visualized as follows (see also refs 9, 10). The initial bleaching of the laser material converts it into very hot ionized plasma. This plasma expands rapidly until the electron densities in the metal wires and the surrounding plastic matrix are equalized. The ratio of materials in the wire and plastic is chosen to provide appropriate lasing conditions after the expansion, when the metallic plasma is left at a lower temperature and density but still in the form of a completely ionized gas. Collisional and radiative recombination produces hydrogenic excited states (states consisting of a single electron bound to the zinc nucleus). Figure 2 shows the associated level diagram and lifetimes of the states. The recombined electrons cascade down through the levels with some of them going through lasing transitions. The pump flux then returns electrons to the continuum by photoionization and the process repeats.

The lifetime for allowed spontaneous emission is $\tau = 1 \times 10^{-15} \lambda^2$ s, where λ is in angstroms. The lower, more energetic transitions will thus have substantially faster decays and correspondingly smaller equilibrium populations, thus potentially providing conditions for population inversion and lasing.

A precise calculation for the system would have to take into account the interplay between the plasma densities and temperatures determining the atomic recombination rates and line broadening,

and the energy- and time-evolution of the pump flux causing reionization and repopulation of higher levels by photoabsorption. Such a calculation is possible only by sophisticated numerical methods. Qualitatively, the rate of photoexcitation and ionization removing electrons from the $1s$ state is insufficient to maintain a population inversion for transitions to the $1s$ state. Additionally photoabsorption^{9,10} re-populates the p -state populations by means of the $1 \rightarrow 2p$, $3p$, etc, transitions, and precludes building up the requisite population inversions to provide lasing to the $n=2$ level. The first lasing level for which a population inversion appears to be possible is the $3d$ level, corresponding to the possible lasing transitions, $4f \rightarrow 3d$ and $5f \rightarrow 3d$. Bunkin *et al.*⁹ estimate the population inversions to be completely adequate to sustain lasing and expect the $5f \rightarrow 3d$ transition to dominate. The number of laser transitions made by an individual atom could well lie in the range of 10–100.

Table 1 lists the effective lasing transition as the $5f \rightarrow 3d$ transition of the zinc hydrogenic system. The energy associated with this $n=5$ to $n=3$ transition is $13.6Z^2(1/3^2 - 1/5^2)$ eV or 0.87 keV, corresponding to a wavelength of 14.2 Å. Whether or not the lasing material is actually zinc or the lasing transition is indeed $5f \rightarrow 3d$, would not seriously affect the overall characteristics described in this article.

To obtain some measure of the upper limit for the achievable output efficiency, crude estimates can be made of the fractional efficiencies involved in the chain of events required to transform fission power into directed laser energy. The energy transfer from explosion energy into X-ray pump flux is ~ 0.7 . The laser assembly subtends ~ 0.7 of the total 4π solid angle into which the pump flux is emitted. The energy of the emitted photons, 0.87 keV, is ~ 0.07 of the energy required to photoionize an atom. The fraction of recombination electrons that pass through the lasing transition level is ~ 0.5 . The effective output efficiency for the lasing element will be ~ 0.2 , due both to scattering out of beam from the lasing element by Compton scattering and to not all of the lasing elements being fully saturated. In addition, inefficiencies of ~ 0.5 will arise from imperfect matching of the energy spectrum and time distribution of the pump pulse to lasing requirements. The combination of these factors gives an estimate for the overall energy conversion of $\sim 0.17\%$. This value is comparable with Chapline and Wood's 'rule of thumb' of $\sim 0.1\%$ for the conversion efficiency for X-ray lasers. The value used in Table 1 is 0.1%.

Lower limit to the output wavelength. The maximum atomic number Z for the lasing material that can be used follows from the requirement that the pump flux of X rays

should be able to ionize the lasing material efficiently. As described earlier, the pump flux has an approximately black-body spectrum with a mean energy of 10 keV. Energies in excess of 20 keV are required to photoionize the K-shell of elements with $Z > 40$, hence such elements will not be an efficient match to the pump flux. An atomic number of 40 would therefore seem to be an upper limit for the lasing material.

From the preceding section it is probable that the most energetic transition available is the $5f \rightarrow 3d$. For a Z of 40 this corresponds to a wavelength of 8 Å, which thus appears to be a lower limit to wavelengths that could be produced in this type of laser.

Beam collimation and focusing (see also ref. 11). Three elements are essential to the attainment of greatest destructive power with this projected weaponry. First, there must be efficient transfer of pump power to laser power; second, the power must be highly collimated or directed; and third, the beams must be aimed with the requisite precision. In this section I examine the theoretical limitations set by considerations of geometrical optics and diffraction.

Diffraction limits the achievable half-angular collimation θ_{diff} to $\approx \lambda/D$ where D is the wire diameter and λ is the wavelength.

Without focusing, the best achievable geometric angular definition is $\theta_{\text{geom}} \approx D/L$, where L is the length of the laser element.

Taking into account both geometric and diffraction effects, without focusing, the optimum value θ_{min} occurs for¹² $D \approx \sqrt{\lambda L}$, so that $\theta_{\text{min}} \approx \sqrt{\lambda/L}$.

For 2-m laser elements the best wire diameter for optimal collimation would be $\sim 5.5 \times 10^{-3}$ cm, and the associated half angle would be $\sim 2.6 \times 10^{-5}$ radians.

However, at the wavelengths of interest the ionized solid plasma, into which material is converted by the pump X-ray flux, has a refractive index n given by $n = 1 - \lambda^2/2\lambda_p^2$, where n is the refractive index, λ is the laser wavelength and λ_p is the plasma wavelength of the medium, which for a non-degenerate plasma is given by $\lambda_p \approx 2\pi\sqrt{\epsilon_0 m_e / Ne^2}$ (mks units), where ϵ_0 is the permittivity constant, m_e the electron mass, e the electric charge and N the number density of electrons in the material. λ_p is ~ 600 Å, for a fully ionized zinc plasma of density 1.5 g cm^{-3} , giving $1 - n \sim 2.5 \times 10^{-4}$ for an incident wavelength of 14 Å.

Conventional lens elements are constructed from shaped interfaces at the boundaries of materials with different refractive indices. Refractive index changes of $\sim 10^{-4}$, at interfaces between high- and low- Z plasmas, allow for the production of lens systems, at the output end of the laser wires, with focal lengths of ~ 2 m. (The radii of curvature (r) of the

faces of a symmetric lens of focal length f are $2f\Delta n$, where Δn is the refractive index change at the boundary. A radius of curvature of 4×10^{-2} cm with $\Delta n = 10^{-4}$ forms a 2-m focal length symmetric lens. This radius of curvature is larger than the diameter of the lasing wire and is thus physically achievable.)

In principle, to eliminate geometrical optics effects the lasing wires could be cones, as shown in Fig. 1c, with an output lens system used to remove geometrical beam divergences. In practice, a variety of methods taking advantage of the expansion characteristics of the plasma and/or using non-uniform material assemblies could be used to achieve focusing by means of refractive index changes in the plasma.

The theoretically achievable collimation, using focusing, is limited by diffraction and the spatial extent of the distribution in origins of the primary photons associated with lasing. This spatial extent is $\sim L/\alpha$, where L is the length and α the amplification factor for the lasing elements. The best achievable collimation with focusing corresponds to a diameter $D \approx \sqrt{\alpha \lambda L}$ at the output end of the lasing element, and $\theta_{\text{min}} \approx \sqrt{\lambda/\alpha L}$.

This corresponds, for $\alpha \approx 25$, to an improvement of a factor of ~ 5 over the limit obtained without focusing, and leads to a lower limit for θ_{min} of 5×10^{-6} radians. Comparable limits were obtained from considerations of the non-uniform mechanical displacements occurring in the lasing elements over the pump times of 10^{-8} s.

The value given in Table 1 for the solid angle containing the output laser beam is $\Delta\Omega \approx \pi(6 \times 10^{-6})(6 \times 10^{-6}) = 1 \times 10^{-10}$ steradians. If focusing proves impossible, the smallest achievable solid angle would be set by the limit for combined geometrical and diffraction effects, corresponding to $\pi \times (2.6 \times 10^{-5})^2$ steradians. This solid angle is 25 times larger than the value used in Table 1 and would result in energy fluxes per unit solid angle lower by a factor of ~ 25 .

Performance testing. Underground testing of nuclear devices inevitably takes place in confined spaces. Thus it might not appear possible, with underground testing, to establish the properties of laser beams of enormous power, beams that instantly destroy all objects they encounter.

A simple method of testing performance would be provided by masking off all but a few of the beams from the laser wires and observing the penetration of these beams into an inert gas several hundred metres away from the laser. The passage of the beam through the gas would be marked by the production of a highly visible hot plasma core, with the depth of penetration providing a measure of the energy intensity. The plasma core would be observed from the side, with micro-

seconds available for recording and transmitting the information to ground level.

Undoubtedly, several experiments can be conducted in parallel using the same nuclear explosion, and after the experience gained from years of underground testing many highly sophisticated test procedures must be available to measure performance parameters, including the laser output energy intensity per unit solid angle. Even so, given the harsh environment of an underground nuclear test, it is not surprising that problems of calibration remain unsolved¹³.

Summary. One can conclude that no fundamental limitations prohibit the development, using currently available technology, of nuclear-pumped X-ray lasers that can instantaneously destroy or incapacitate objects in space at ranges in excess of 2,000 km. Obvious problems are associated with the choice of lasing materials and of the desired spectroscopic lasing transition, the configuration and power levels of the fission pump source, and mechanical assembly and the need for precise alignment. Equally critical are questions of whether collimation can be substantially enhanced by focusing and whether alignment tolerances can be maintained in the hot, high-pressure plasma into which the lasing system is transformed.

It is clear that very complex and possibly conflicting trade-offs enter into the engineering, design and construction of such weaponry, and that neither the time scale nor final performance levels can be predicted at present¹. But it is also clear that, in the absence of agreements to regulate research or deployment, over the next decades this weaponry could be successfully developed for operational use and enter the arsenals of one or both superpowers.

I thank W. Kirk, H. Lynch, and L. Rosenberg for discussions and helpful comments.

1. Robinson, C. Jr. *Aviat. Week*, 23 February, 25 (1981).
2. *Report of the Study Group on the Science and Technology of Directed Energy Weapons* (American Institute of Physics, New York, 1987).
3. Carter, A. B. "Directed Energy Missile Defense in Space—A Background Paper" OTA-BP-ISC-26 Office of Technology Assessment, Washington, DC, 1984).
4. Bethe, H. & Garwin, R. L. *Daedalus* 114, 53–71 (1985).
5. Glasstone, S. & Dolan, P. *The Effects of Nuclear Weapons*, 3rd Edn (US Department of Defense and the US Department of Energy, Washington, DC, 1977).
6. Slater, J. C. *Quantum Theory of Atomic Structure* (McGraw Hill, New York, 1960).
7. Chapline, G. & Wood, L. "Conditions for X-ray Laser Action" Nat. Tech. Inf. Serv. UCRL-75184-Rev II (US Department of Commerce, Washington, DC).
8. Wood, L. & Chapline, G. *Physics Today* 6, 40 (1975).
9. Bunkin, F., Derzhiev, V. & Yakovlenko, S. *Sov. J. Quant. Electron.* 11, 971 (1981).
10. Boiko, V. A., Bunkin, F. V., Derzhiev, V. I. & Yakovlenko, S. *IEEE J. Quant. Electron.* 3, QE 20 (1984).
11. Feynman, R. et al. *Lectures on Physics* (Addison-Wesley, Reading, Massachusetts, 1963–1965).
12. Walbridge, E. *Nature* 310, 180 (1984).
13. Jeffrey-Smith, R. *Science* 230, 646 (1985).

D. M. Ritson is in the Department of Physics, Stanford University, Stanford, California 94305, USA.

Geological and industrial implications of extensive-dilatancy anisotropy

Stuart Crampin

British Geological Survey, Murchison House, West Mains Road, Edinburgh EH9 3LA, UK

Extensive-dilatancy anisotropy is the hypothesized distribution of stress-aligned fluid-filled microcracks pervading most rocks in the Earth's crust. The geometry of the cracks and the aligning stress-field can be monitored by analysing the waveforms of shear waves propagating through the rockmass. The ability to estimate some of the parameters of the crack and stress geometry by analysing shear-wave particle displacements has widespread implications.

AN 'anisotropic' solid contains an internal structure such as aligned crystals or aligned cracks so that the elastic behaviour varies with direction. This contrasts with a uniform 'isotropic' solid where the properties are the same in all directions. Such anisotropy in rocks in the Earth would have two major effects on seismic wave propagation. First, the velocities of both P-waves (with longitudinal particle motion) and shear-waves (with transverse motion) vary with the direction of propagation. Second, on entering a region of anisotropy, shear-waves split into the two (or more) phases with fixed velocities and fixed polarizations that propagate in that particular direction through the anisotropy.

Until recently, most seismic investigations in the Earth's crust were confined to P-waves recorded on vertical-component instruments. It was found that layers and blocks of isotropic rock can model P-wave arrival times with considerable accuracy. The fact that many rocks are crystalline and anisotropic on a small scale was interpreted as indicating that the orientation of the anisotropy was sufficiently random at larger scales that the comparatively long-wavelength seismic waves sample only the average (isotropic) properties of the rockmass¹. In addition, it is difficult to recognize smooth variations of P-wave or shear-wave velocities unless arrival times can be observed over a wide range of directions in a single homogeneous layer, and this is seldom possible in the usually complicated crustal structure. Thus shear-wave splitting is likely to be the most reliable indicator of anisotropy in the crust. It is only in the last three or four years that digital recording and analysis technology has advanced sufficiently for shear-wave splitting to be easily recorded and identified in short-period shear-waves.

Shear-wave splitting, also known as birefringence and double-refraction, is diagnostic of some form of effective anisotropy along the raypath^{2,3}. Following the recognition that such behaviour would be expected in propagation through aligned microcracks, it was suggested² that it might be possible to monitor earthquake dilatancy by analysing shear waves recorded above small earthquakes. (Dilatancy is the increase in volume as microcracks open before failure in rock specimens subjected to laboratory stresses greater than half the eventual fracture strength of the intact sample⁴.)

Observational evidence

To try to detect such shear-wave splitting, closely spaced networks of three-component seismometers were deployed above a swarm of small earthquakes near the North Anatolian Fault (NAF) in Turkey. These Turkish Dilatancy Projects operated for a few months in 1979 (TDP1), a few months in 1980 (TDP2), and for six months in 1984 (TDP3—currently being analysed). Shear-wave splitting was identified⁵⁻⁷ in all shear

wavetrains recorded within the shear-wave window at the surface (the shear-wave window is defined by an angle of incidence less than the critical angle $\sin^{-1} V_s/V_p$, approximately 35° for Poisson's ratio of 0.25^{8,9}). The shear waves were scattered (this is now recognized as being caused principally by interactions with the irregular surface topography) but displayed remarkably consistent alignment of the shear-wave polarizations. The polarizations of the leading (faster) split shear-waves were parallel to the direction of maximum horizontal compression along that particular section of the NAF¹⁰. The parallel orientations and several other features of the records are consistent with propagation through the effective anisotropy of uniform distributions of parallel vertical water-filled microcracks striking perpendicular to the (horizontal) minimum compressional stress¹⁰. The seismograms indicate anisotropy (the crack distribution) throughout the region surrounding the swarm, not just near the stress-concentrations immediately adjacent to the focus¹⁰.

Shear-wave splitting above small earthquakes and parallel polarizations have now been widely observed in many places around the world (Table 1), and there are no counter examples where short-period shear-waves recorded subsurface or within the shear-wave window at the surface do not show evidence of shear-wave splitting. Particularly important observations are that temporal variations in the shear-wave splitting have been recognized in two seismic gaps where large earthquakes are expected^{11,12}. Shear-wave splitting has also been observed above acoustic events in geothermal experiments^{13,14}, in shear-wave vertical-seismic-profiles (VSPs) in a geothermal reservoir^{15,16}, in sedimentary basins¹⁷⁻²⁰, and elsewhere^{21,22}, and in most shear-wave reflection surveys made for exploration and production purposes in hydrocarbon reservoirs^{23,24}. The polarizations of the shear-waves in the geothermal reservoir^{15,16} are directly related to visible macro-fractures, and the oil company reflection surveys and VSPs are interpreted as indicating the orientation of subsurface fractures^{19,20,23,24}.

Shear-wave splitting

The many complications of the shear wavetrain were attributed in the past to scattering at (unspecified) heterogeneities. Recent observations of shear-wave splitting and the remarkable parallelism of the polarizations clearly indicate an internal structure (anisotropy) and not a random distribution of scatterers. Any single observation of shear-wave splitting can usually be simulated by propagation through a combination of isotropic discontinuities. However, the almost universal presence of shear-wave splitting and the uniform alignment of the leading shear-waves normal to the direction of minimum compressional stress cannot be explained by any realistic combination of isotropic

Table 1 Seismic observations of anisotropy in the crust

Location	Rock type	Earthquake mechanisms	Shear-wave alignments		References
			Station	Network	
Shear-wave splitting above small earthquakes:					
North Anatolian Fault, Turkey	Sediments, metamorphic	Strike-slip	Parallel	Parallel	5, 6, 10
Peter the First Range, USSR	Poorly consolidated sediments	Thrust	Parallel	Parallel	27
North Wales, UK	Granite, slate, metamorphic	Strike-slip	Parallel	Parallel	45
Kinki District, Japan	Granite, slate, sediments	Strike-slip	Parallel (granite, sediments), scattered (slate)	Parallel (granite, sediments), scattered (slate)	42
Kyushu, Japan	Various	Normal	Parallel	Various	46
Quebec, Canada	Sediments	Strike-slip	Parallel	Parallel	47
Anza, California, USA	Various	Strike-slip	Parallel	Various	11
Shear-wave splitting above shallow events in geothermal reservoirs:					
Takinoue, Japan	Metamorphic	—	Parallel	Parallel	13
Cornwall, UK	Granite	—	Parallel	Parallel	14
Shear-wave splitting in VSPs:					
Paris Basin, France	Sandstone, shale	—	*	*	17, 18
Texas, USA	Austin chalk	—	*	—	19, 20
Four sites in California	Various	—	*	—	15, 16, 21, 22
Shear-wave splitting in reflection surveys:					
Twelve sites across North America	Various	—	*	—	23, 24
Velocity anisotropy in refracted P-waves					
Mount Hood, Oregon, USA	Metamorphic	—	—	—	35

* Impulsive shear-waves arrivals from earthquakes usually possess easily identifiable polarizations. Emergent harmonic arrivals from shear-wave vibrators can only be interpreted accurately by modelling with synthetic seismograms.

structures^{25,26}, and the uniformity of the observations in Table 1 suggests that some form of effective anisotropy is present in most crustal rocks.

In seeking the cause of this seismic anisotropy, three phenomena appear to be critical: (i) The almost universal presence of anisotropy in many types of rocks at many depths and in many geological and tectonic environments^{10,14,23,24}. (ii) The uniformity of the alignment of the polarization of the leading split shear-waves^{6,11,14,27} normal to the current minimum horizontal compression. And (iii) the temporal variations of the behaviour of shear-wave splitting in seismic gaps^{11,12}.

The explanation for these phenomena must be sought among the five major causes of seismic anisotropy in the Earth identified in a recent review²⁸.

Aligned crystals. Crystals may be aligned by stress by creep and re-crystallization. This can only occur at temperatures greater than those usually found in the relatively cool upper crust where splitting is observed^{6,14,27}. Splitting also occurs in poorly consolidated and contorted sedimentary sequences where uniformly aligned crystals are unlikely to be present²⁷.

Lithologic anisotropy caused by a fabric such as aligned grains. Any lithologic anisotropy is likely to be confined to particular rock strata, will show little time variation, and cannot satisfy (i), (ii) and (iii) above.

Long-wavelength anisotropy such as that caused by periodic thin layers. Again, such anisotropy is likely to be limited in space and fixed in time, and cannot satisfy the three critical observations.

Direct stress-induced anisotropy. Direct stress-induced anisotropy only becomes significant when the stress becomes a large proportion of the eventual fracture strength. This could be important at stress concentrations very close to the extremities of faults, but the stress in stable areas where splitting has also been observed is probably several orders of magnitude too small to cause observable effects²⁹. In any case, rock physics experiments demonstrate that rocks deform by dilatancy (opening of

microcracks) before direct stress-induced anisotropy becomes significant⁴.

Crack-induced anisotropy caused by aligned cracks and microcracks. The next sections argue that cracks and particularly microcracks are universally present in the crust, are aligned by the current stress-field and, most important, have configurations which can be easily modified by changes in stress on a short enough time-scale to cause the observed temporal variations.

Extensive-dilatancy anisotropy

Distributions of stress-aligned fluid-filled microcracks must be expected in the crust³⁰. Every sedimentary rock retains water in its interstices after deposition, and the long enduring regional stress-fields will tend to align the pore space normal to the minimum compressional stress by such processes as subcritical crack-growth³¹. This will not result in perfectly flat parallel microcracks but it may cause sufficient flattening of the pore space to be anisotropic to shear-waves as the observed anisotropy is relatively weak. Such effects have been demonstrated in rock physics experiments³² (but at larger stresses in order to obtain results in an acceptable time scale). Similarly, prograde metamorphic processes release chemically bound water from most igneous rocks³⁰ and the only mechanism for such water to be released is by hydraulic fracture³⁰ at high pore-fluid pressures into inter- and intra-granular microcracks which will again be aligned normal to the minimum compressional stress. Thus distributions of cracks and (principally) microcracks must exist within the *in situ* rockmass, and although their geometry can be modified by stress, they are not directly stress-induced. This phenomenon of stress-aligned cracks, originally recognized above earthquake foci, is known as extensive-dilatancy anisotropy (EDA)²⁵, and has now been recognized throughout much of the crust^{26,33}. Typical observations of shear-wave splitting suggest a maximum differential velocity-anisotropy of about 4% yielding a crack density of $N a^3 / V = 0.04$ where N is the number of cracks of radius a in a volume V .

This is equivalent to a crack of diameter 0.7 in each unit cube.

One of the fundamental difficulties of testing this hypothesis of EDA cracks is that *in situ* microcracks in the crust are essentially inaccessible to physical examination. Table 2 lists twenty largely independent phenomena controlling the behaviour of (isolated) fluid-filled microcracks at depth in the crust^{26,33}. The effects of almost all of these phenomena are directly or indirectly controlled by stress. This means that whenever we approach *in situ* microcracks by drilling, mining, or excavating, the rock is disturbed by partial de-stressing, and a new stress anomaly imposed. Since isolated fluid-filled microcracks are probably a comparatively mobile phenomenon in the crust, the *in situ* microcrack geometry is usually destroyed. Even well-logs, which may appear to be examining *in situ* conditions, are sampling material which has the original stress-field severely modified and the *in situ* microcrack geometry is disturbed within a metre or two of the wellbore. *In situ* microcracks are likely to leave residual traces in the rock, but because EDA cracks are frequently isolated microcracks in igneous and metamorphic rocks and are pores with (usually) weak preferred orientations in sedimentary rocks, there are also considerable difficulties in examining traces of EDA in the rock physics laboratory. Extrapolating from conditions in the rock physics laboratory to *in situ* conditions requires extreme caution and, as Table 2 suggests, requires a multi-variable extrapolation over many orders of magnitude, only a few of which can be monitored at any one time in the laboratory. This inaccessibility means that seismic shear-waves are the principal technique for monitoring crack properties *in situ* and the exact configuration of EDA cracks can only be estimated.

Since the wavelengths of shear-waves are many times the likely crack dimensions, shear-waves can give us little information about the crack dimensions. However, the dimensions of EDA cracks are likely to span several orders of magnitude. In intact igneous and metamorphic rocks, the dimensions may vary from submicrometre to a few tens of micrometres and are initially isolated with few connections to other cracks. In sedimentary rocks, the EDA cracks may be larger, up to 1 or 2 mm in some rocks, and may be merely the oriented portion of a randomly oriented pore-space of much greater porosity. In fractured beds the crack dimensions may be several metres. We use the term 'EDA cracks' to include this wide range of phenomena. All EDA cracks are likely to have high pore-fluid pressures equalling the lithostatic stress but possible very low porosities particularly in igneous and metamorphic rocks.

Shear-waves are very sensitive to the interaction with the free surface and although the general character of the patterns of polarization can be modelled with synthetic seismograms¹⁰, exact matching of observations and synthetics at the surface is unlikely to be possible except in exceptionally simple structures. This means that subsurface recordings in VSPs are going to be particularly important for determining the properties of EDA cracks.

The only set of three-component shear-wave VSPs to be modelled by synthetics at this time (in the sedimentary sequences of the Paris Basin^{17,18}) provides the first quantitative evaluation of EDA cracks. The detailed waveforms of shear waves recorded downwell from shear-wave vibrators from a wide range of offsets on the surface are well-matched by synthetic seismograms of shear waves propagating through a very simple two-layered model containing a weak distribution of parallel liquid-filled microcracks (crack densities of 0.03 and 0.01, respectively). The microcracks strike parallel to the direction of horizontal compressional stress^{17,18}.

A similar quantitative evaluation was also obtained from azimuthal velocity variations of P-waves in a refraction survey around Mount Hood, Oregon³⁴. The velocity variations were interpreted in terms of fluid-filled microcracks, which because of the high heat-flow around Mount Hood may indicate the

Table 2 Phenomena controlling the behaviour of *in situ* cracks

External conditions:

- | | |
|------------------------|----------------------------|
| (1) Lithostatic stress | (2) Deviatoric stress |
| (3) Temperature | (4) Properties of rockmass |

Internal conditions:

- | | |
|---------------------------------------|---|
| (5) Pore-fluid pressure | (6) Compressibility of pore-fluid |
| (7) Viscosity of pore-fluid | (8) Debris in crack void |
| (9) Vapour/liquid ratio in pore-fluid | (10) Properties of pore fluid under high temperatures and pressures |

Dynamic conditions:

- | | |
|---------------------------|----------------------------|
| (11) Rate of strain | (12) Rate of crack healing |
| (13) Rate of crack growth | |

Crack parameters:

- | | |
|---|--|
| (14) Orientation | (15) Dimensions |
| (16) Aspect ratio | (17) Distribution |
| (18) Smoothness of crack faces | (19) Connectedness (degree of isolation) |
| (20) Geometry (parallel, biplanar, etc) | |

presence of fluid-filled microcracks throughout most of the crust³⁵. It should be noted that such P-wave studies are limited to studies of horizontally propagating head-waves, and are unlikely to be repeated very frequently as the survey around Mount Hood was exceptionally comprehensive with hundreds of source to geophone paths in a comparatively small area.

Originally proposed as existing only around earthquake foci²⁵, there is now a large amount of evidence (Table 1) that suggest that EDA cracks exist in many (perhaps most) rocks in the crust^{26,33}, and shear-wave splitting as an indicator of subsurface fracture alignments is now being investigated by the hydrocarbon industry^{17-20,23,24}. Seismic evidence in the papers cited in Table 1 suggest that most rocks in at least the upper 10 to 20 km of the brittle crust contain EDA cracks. The cracks are typically aligned vertically and perpendicular to the (usually horizontal) minimum compressional stress. Recognition that distributions of EDA cracks exist in most rocks is, I suggest, a fundamental advance in our understanding of the physical state of *in situ* rock which has many geological and industrial implications^{26,33}.

Geological implications

The presence of fluids at all levels in the Earth's crust is well established³⁰, although the mechanism by which fluids are retained at necessarily high pore-fluid pressures at depths in the crust has not previously been specified. Fyfe *et al.*³⁰ suggest, for example, that most fluids are retained in cracks (veins) and fractures, and the implication that metamorphic hydraulic fracture releases connate water from most grains into (initially isolated) microcracks throughout the rockmass is not followed through. The evidence referred to above suggests that many rocks in the crust contain an internal structure of fluid-filled EDA microcracks aligned by the contemporary regional stress field. EDA cracks provide a mechanism for the presence of fluids in the crust, and are, I suggest, the primary cause of a wide variety of previously unrelated phenomena:

High pore-fluid pressures necessary for fault movements. Many tectonic processes, particularly thrust faulting, require high pore-fluid pressures to reduce the normal stresses sufficiently to allow slippage to occur. Fluids from even low porosity EDA cracks will be concentrated and channelled by subcritical crack-growth at high-stress concentrations into regions of higher dilatancy near the eventual fault zone. The fault planes will be lubricated, frictional forces reduced, and slippage may occur.

Flexibility of crustal plates. The presence of EDA cracks allows large crustal plates and blocks to deform to accommodate small changes of strain imposed by plate tectonics and fault movements. This tends to reduce earthquake activity so that relatively aseismic areas may exist within plates and blocks moving over

the surface of a non-spherical mantle and undergoing deformation around their margins.

Main driving mechanism for earthquake precursors. Modifications and manipulations of EDA cracks by stress changes are the common driving mechanisms for the large variety of precursors that are intermittently observed before earthquakes. Since EDA cracks are necessarily aligned and their effects anisotropic, each precursory phenomenon will depend on the geometry of the particular earthquake stress-field, the geological structure, and the geometry of the recording sites². Thus, the search for particular precursors may be difficult and lead to unpredictable results.

Relative P- and S-wave attenuation. The theoretical attenuation due to scattering at crack margins of shear-waves propagating through liquid-filled cracks is greater than the theoretical attenuation of P-waves by more than three times (sometimes much more, depending on direction)³⁶. This is approximately the relative attenuation observed in the Earth. The attenuation caused by propagation through liquid-filled cracks is not well understood. Although the effects of scattering at crack margins are too small, the introduction of a small gas bubble or turbulent flow could increase the attenuation by orders of magnitude. Thus EDA cracks are certainly one of the causes for the greater shear-wave attenuation in the Earth, and the variation of attenuation with azimuth and incidence angle is the reason for some of the inconsistencies in observations of attenuation in the crust. It is worth noting that because distributions of microcracks change the effective elastic constants of the matrix, the microcracks may affect a wide range of seismic wavelengths.

Relative terrestrial and lunar attenuation. The marked difference in attenuation between the Earth and the Moon is usually attributed to the lack of evaporites on the Moon³⁷. In the Earth's crust (and by implication, the mantle, since isolated inclusions of melt or other fluids in the mantle are also not going to be easily closed by lithostatic stress), the attenuation is principally the result of scattering by fluid-filled EDA cracks. In the Moon, where there are no pore-fluid pressures to keep cracks open, cracks will be closed or absent and there will be little attenuation.

High conductivity deep in the crust. The electrical conductivity at all levels in the crust is anomalously high by several orders of magnitude relative to dry rocks in the laboratory³⁸ and may vary with azimuth by an order of magnitude. The high conductivity and the variation with azimuth are the result of aligned fluid-filled EDA cracks both at shallow depths, as measured by well logs, and at deeper depths (down to 20 km) as determined by large-aperture long-period measurements at the surface.

Earthquake prediction research

The variety of precursors, which are sporadically observed before earthquakes, are usually ascribed to the indirect effects of the variation of stress. The most direct effects of changes of stress are modifications of the configurations of the EDA cracks in the stressed rockmass. Since the behaviour of shear waves is controlled by three-dimensional variations in the crack geometry, it is probable that analyses of shear-wave splitting could lead to direct determinations of changes in the crack configuration and hence stress changes, and thus offer a technique for earthquake prediction. No larger event has yet been "captured" within any of the small number of existing three-component seismometer networks and no definite changes in the configuration of EDA cracks has yet been recognized, although some possible temporal changes may have been identified (see the paragraph referring to aspect ratio, below).

There are several ways in which changes of stress may alter the configuration of EDA cracks:

Changes in orientation of the cracks. Changes in the direction of stress are likely to alter the orientation of EDA cracks. This would result in changes in the shear-wave polarizations, which are the most easily identifiable and stable parameters of shear-

wave splitting. However, stress changes before an earthquake are usually confined to modifying the magnitude of the stress but preserving the same principal axes. This would not alter crack orientations and there would be no change in shear-wave polarizations. Crampin and Booth *et al.* observed possible changes in the polarization direction near one of the TDP sites in Turkey between 1979 and 1980, and suggested that it might have been caused by local stress polarization-anomalies (refs 6 and 10). However, the same polarizations are seen to persist in TDP3 and the anomalies are now thought to be caused by the interactions of the shear-wave arrivals with the very steep local topography.

Changes in crack density. Increasing stress is likely to increase crack densities by promoting crack growth (by subcritical crack growth³¹) rather than by opening new cracks, which occurs only at high stress⁴. Increasing crack density would increase the differential shear-wave anisotropy and increase the time delay between the split shear-waves. This should be recognizable if exactly repeatable sources of shear waves can be monitored with the same instrumental networks (see below).

Changes of crack aspect-ratio. The direct elastic effect of increasing stress on a rock containing EDA cracks is likely to increase the aspect ratio (similar to the increasing conventional high-stress dilatancy in rock specimens subject to uniaxial compression). This would have most effect on modifying the time delays between shear waves propagating at between 40° and 75° to the crack normal^{35,36}. Peacock *et al.*¹¹ have reported temporal changes in shear-wave splitting at a seismic station near the Anza seismic gap in California that could be interpreted as stress-induced increases of the aspect ratio of EDA cracks during the build-up of stress before the expected large earthquake. If this is the first observation of a precursory variation in shear-wave splitting, it could be a very important breakthrough in earthquake prediction research.

Changes of pore-fluid content. The dilatancy-diffusion theory of earthquake precursors suggested that changes of stress could cause fluids to drain from dilated cracks modifying wave propagation through the cracked rock³⁹. It is suggested here that this phenomenon is unlikely to occur for small often-isolated EDA cracks except at high stress concentrations in a limited volume in the immediate vicinity of a fault (but see the next paragraph). The limited volume is likely to be too small to affect shear waves significantly.

Changes of pore-fluid pressure. Changing the dimensions of cracks by crack growth or changing aspect-ratio is likely to alter the pore-fluid pressure (at least temporarily). Significant increases of aspect-ratio may lead to under-saturation and the introduction of small vapour-filled bubbles into the cracks. Variations of pore-fluid pressure and particularly the introduction of small bubbles have little direct effect of seismic velocities and are unlikely to be easily recognized in patterns of particle displacement. The largest effect of the introduction of bubbles will be to increase attenuation. Attenuation, however, is difficult to monitor from earthquake sources, but changes in attenuation should be easily recognized if experiments can be repeated with identical sources (see below).

The principal effect of changes in aspect ratio, the parameter most likely to be modified by changes of stress, is that the time delay between the split shear-waves will be modified. Since shear waves have severe interactions with the free surface^{8,9}, and small earthquakes have unrepeatable source radiation⁴⁰, it is unlikely that time delays can be accurately monitored using surface recordings of natural events. However, three-component VSPs using repeated shear-wave sources are likely to be very sensitive to even small changes in time delays, and repeated shear-wave VSPs is the preferred technique for monitoring small changes in the stress field and hence for predicting earthquakes.

Shear-wave VSPs would also allow changes in attenuation to be monitored. Many of the changes discussed above, particularly increases in aspect ratio, are expected to modify the

shear-wave attenuation. This has received little attention in the past because all existing investigations have used earthquake sources which are too arbitrary to enable attenuation to be monitored.

Since modifying EDA cracks is the most direct effect of changes of stress before earthquakes, and since analyzing EDA cracks is the only way that the essentially anisotropic nature of all stress-induced effects can be explored, monitoring shear-wave splitting with shear-wave VSPs could be a very promising technique for earthquake prediction research.

Industrial implications

Cracks and stress are frequently crucially important in drilling, mining, or excavating, and the ability to monitor the geometry of cracks and stress by analysis of, particularly, shear-wave VSPs has a number of possibly very important applications. The following list is not exhaustive.

Predicting orientations of hydraulic fractures. The orientations of stress-aligned EDA cracks, which may be estimated by analysing the waveforms of shear-waves, predict the orientations of any subsequent hydraulic fractures. This has been demonstrated at the hot-dry-rock geothermal experiment in Cornwall¹⁴. The polarizations of shear-waves recorded at the surface from a very small acoustic event, stimulated when testing the hydraulic equipment, had the same polarizations as those from larger events when the fracture system was established and both were oriented parallel to the strike of the major hydraulic fractures¹⁴. This demonstrates that the internal structure of aligned EDA cracks in the rockmass existed before the structure of major hydraulic fractures was established, and that both crack geometries had the same anisotropic symmetry. Modelling the detailed waveforms of the shear-waves with synthetic seismograms¹⁸ demonstrates that three-component shear-wave VSPs are likely to be a more accurate technique for estimating crack and stress orientations than monitoring recordings made at the surface.

Estimating internal structure of hydrocarbon reservoirs. The analysis of shear-wave VSPs provide direct geophysical information about the *in situ* reservoir rock, which can be related to preferred directions of flow and other properties of the reservoir. Detailed modelling of shear waveforms in three-component shear-wave VSPs¹⁸ yields accurate estimates of the internal structure throughout the rockmass surrounding the well, not just in the immediate vicinity of the well hole which may be anomalous. This is a new source of directly interpretable information which could be important for the appraisal and evaluation of hydrocarbon reservoirs and for optimizing production strategies for secondary and tertiary recovery.

Investigating geometry of fractures in geothermal reservoirs. Shear waveforms, particularly in VSPs, can yield estimates of orientations of *in situ* microcracks and macrofractures in both wet and hot-dry-rock geothermal reservoirs^{14,15}, although at present none have been fully interpreted. Recording the wavetrains of acoustic events on three-component geophones downwell in the centre of the cracked reservoir is likely to be particularly informative, since the behaviour of the wavetrains is controlled by the separation of the major hydraulic fractures and the penetration of the cooling cracks normal to the major crack faces.

Investigating nuclear waste depositories in solid rock. Estimating the orientations of EDA cracks by VSPs before nuclear waste disposal in a repository in intact rock (possibly clay) will indicate the orientations of eventual thermal-expansion or leakage cracks and allow the position of the depository to be optimized. Similarly, repeated VSPs after deposition should allow the extent of the cracking to be monitored.

Investigating rockbursts in mines. Rockbursts are explosive breakouts of walls or roofs which occur in some mines and may cause extensive loss of life and equipment. Shear-wave splitting caused by EDA cracks has been observed in mines (ref. 41 and L. O. Nicolaysen, personal communication). Investigating such

cracks by monitoring shear waves is likely to demonstrate the detailed stress variations prior to rockbursts. The particular importance to earthquake and rockburst prediction research is that the source of rock bursts can be dissected after the event so that the relics of the crack behaviour at the source can be examined (L. O. Nicolaysen, personal communication). This could lead to important advances in understanding source processes.

Estimating slope stability. EDA cracks appear to exist in all types of rock from poorly consolidated sediments²⁷ to hard igneous rock^{6,14,42}. EDA cracks in unstable slopes are likely to display larger crack densities (probably by crack growth rather than by new cracks opening) and larger aspect ratios as the instability increases. Such changes in crack configurations could be monitored by analysing shear-waves in repeated cross-hole or VSP experiments using shear-wave sources.

Estimated overburden fractures in open-cast mining. It is necessary to obtain estimates of overburden fracture density in order to optimize the use of expensive mining equipment in open-cast mining⁴³. Recording shear-waves with three-component geophones in cross-hole shooting, or VSPs as appropriate, should yield detailed estimates of crack geometry and crack density.

Discussion

The previous sections suggest some of the possible implications and applications of monitoring EDA cracks with shear waves. Estimating the internal structure (fracture orientations) of reservoirs is already being investigated by the hydrocarbon industry^{19,20,23,24}. Since shear-wavetrains contain three or four times more information about the structure along the raypath and the source than the equivalent P-wavetrain⁴⁴, it is clear that monitoring shear-waves has great potential for many earth science investigations.

One of the major reasons why EDA cracks have not been recognized earlier is that the majority of short-period seismic investigations in the Earth's crust have been restricted in the past to analysis of P-wave arrival-times for which single-component vertical seismometers provide adequate recordings. Thin liquid-filled cracks have very little effect on P-wave propagation^{2,3,36} and the isotropic P-wave velocity models, that have been so enormously successful in describing the shallow Earth structure, are generally valid. The internal structure of aligned EDA cracks only becomes apparent if shear-wave splitting is recorded by three-component digital recorders so that polarization diagrams can be monitored. As yet there are comparatively few suitable recordings in either earthquake or exploration seismology, although the number is rapidly increasing now that the importance of anisotropy has been recognized.

It appears that the presence of EDA microcracks and the ability to monitor crack and stress geometry with shear-waves opens new ways of examining *in situ* rocks (and possibly other solids) that is likely to have important implications and applications over a very wide field. The modelling of three-component shear-waves in VSPs in the Paris Basin^{17,18} is the first time that the detailed waveforms of short-period shear-waves have been successfully modelled by synthetic seismograms. Since shear waves contain so much more information than P-waves⁴⁴, the developments reviewed here are potentially very significant.

I thank David C. Booth for his comments on the manuscript. The work was supported by NERC with some indirect support from the USGS.

1. Ewing, W. M., Jardetzky, W. S. & Press, F. *Elastic Waves in Layered Media* (McGraw-Hill, New York, 1957).
2. Crampin, S. *Geophys. J.R. astr. Soc.* **53**, 467-496 (1978).
3. Crampin, S. *Wave Motion* **3**, 343-391 (1981).
4. Brace, W. F., Paulding, B. W. & Scholz, C. *J. geophys. Res.* **71**, 3939-3953 (1966).
5. Crampin, S. *et al. Nature* **286**, 874-877 (1980).
6. Booth, D. C., Crampin, S., Evans, R. & Roberts, G. *Geophys. J.R. astr. Soc.* **83**, 61-73 (1985).
7. Crampin, S., Evans, R. & Uger, S. B. *Geophys. J.R. astr. Soc.* **83**, 1-16 (1985).
8. Evans, R. *Geophys. J.R. astr. Soc.* **76**, 165-172 (1984).
9. Booth, D. C. & Crampin, S. *Geophys. J.R. astr. Soc.* **83**, 31-45 (1985).

10. Crampin, S. & Booth, D. C. *Geophys. J.R. astr. Soc.* **83**, 75-92 (1985).
11. Peacock, S., Crampin, S. & Fletcher, J. B. *J. geophys. Res.* (submitted).
12. Chen, T.-C., Booth, D. C. & Crampin, S. *Eos* **67**, 1117 (1986).
13. Kaneshima, S., Ito, H. & Mitsuhiro, S. *Abstracts, Second Int. Work. Seismic Anisotropy, Moscow, 1986*, 128 (1986).
14. Roberts, G. & Crampin, S. *Int. J. Rock Mech. Min. Sci.* **23**, 291-301 (1986).
15. Majer, E. L., McEvilly, T. V. & Eastwood, F. S. *Eos* **66**, 950 (1985).
16. Daley, T. M., Majer, E. L. & McEvilly, T. V. *Eos* **67**, 1116 (1986).
17. Crampin, S. & Bush, I. *Expanded Abstracts, 56th Ann. Int. SEG Meeting, Houston, 1986*, 481-484 (1986).
18. Crampin, S., Bush, I., Naville, C. & Taylor, D. B. *The Leading Edge* Vol. 5, 35-39 (1986).
19. Johnston, D. H. *Expanded Abstracts, 56th Ann. Int. SEG Meeting, Houston, 1986*, 464-466 (1986).
20. Becker, D. F. & Perelberg, A. I. *Expanded Abstracts, 56th Ann. Int. SEG Meeting, Houston, 1986*, 466-468 (1986).
21. Leary, P. C. & Li, Y.-G. *Eos* **67**, 1116 (1986).
22. Li, Y.-G., Leary, P. C. & Aki, K. *Eos* **67**, 1117 (1986).
23. Alford, R. M. *Expanded Abstracts, 56th Ann. Int. SEG Meeting, Houston 1986*, 476-479 (1986).
24. Willis, H. A., Rethford, C. L. & Bielanski, E. *Expanded Abstracts, 56th Ann. Int. SEG Meeting, Houston, 1986*, 479-481 (1986).
25. Crampin, S., Evans, R. & Atkinson, B. K. *Geophys. J.R. astr. Soc.* **76**, 147-156 (1984).
26. Crampin, S. *First Break* **3**, 12-15 (1985).
27. Crampin, S. *et al. Geophys. J.R. astr. Soc.* **84**, 401-412 (1986).
28. Crampin, S., Chesnokov, E. M. & Hipkin, R. G. *Geophys. J.R. astr. Soc.* **76**, 1-16 (1984).
29. Dahlen, F. A. *Bull. Seism. Soc. Am.* **62**, 1183-1193 (1972).
30. Fyfe, W. S., Price, N. J. & Thompson, A. B. *Fluids in the Earth's Crust* (Elsevier, Amsterdam, 1978).
31. Atkinson, B. K. *J. geophys. Res.* **89**, 4077-4114 (1984).
32. Rai, C. S. & Hanson, K. E. *Expanded Abstracts, 56th Ann. Int. SEG Meeting, Houston, 1986*, 471-473 (1986).
33. Crampin, S. & Atkinson, B. K. *First Break*, **3**, 16-20 (1985).
34. Kohler, W. M., Healy, J. H. & Wegener, S. S. *J. geophys. Res.* **87**, 339-355 (1982).
35. Crampin, S., McGonigle, R. & Ando, M. *J. geophys. Res.* **91**, 12703-12710 (1986).
36. Crampin, S. *Geophys. J.R. astr. Soc.* **76**, 135-145 (1984).
37. Goins, N. R., Dainty, A. M. & Toksöz, M. N. *J. geophys. Res.* **86**, 378-388 (1981).
38. Shankland, T. J. & Ander, M. E. *J. geophys. Res.* **88**, 9475-9484 (1983).
39. Aggarwal, Y. P., Sykes, L. R., Armbruster, J. & Sbar, M. L. *Nature* **241**, 101-104 (1973).
40. Evans, R., Asudeh, I., Crampin, S. & Üçer, S. B. *Geophys. J.R. astr. Soc.* **83**, 47-60 (1985).
41. Young, R. P. in *Proc. 4th Conf. Acoustic Events/Microseismic Activity in Geological Materials* (Transtech, in the press).
42. Kaneshima, S., Ando, M. & Crampin, S. *Phys. Earth. Planet. Int.* **45**, 45-58 (1987).
43. Young, R. P. & Hill, J. J. *Geophysics* **51**, 302-323 (1986).
44. Crampin, S. *Geophysics* **50**, 142-152 (1985).
45. Peacock, S. *Geophys. J.R. astr. Soc.* **81**, 341 (1985).
46. Kaneshima, S., Ito, H. & Mitsuhiro, S. *Abstracts, Second Int. Work. Seismic Anisotropy, Moscow, 1986*, 127 (1986).
47. Buchbinder, G. G. *R. Geophys. Res. Lett.* **121**, 425-428 (1985).

ARTICLES

Rational modification of enzyme catalysis by engineering surface charge

Alan J. Russell & Alan R. Fersht

Department of Chemistry, Imperial College of Science and Technology, London SW7 2AY, UK

Changing the surface charge of subtilisin by site-directed mutagenesis produces enzymes with significantly shifted pH-activity profiles, higher catalytic activities and altered specificities. Insight into water as a dielectric, the role of ions in electrostatic shielding and the field effects on catalysis is obtained and suggests rules for tailoring pH-activity profiles.

ELECTROSTATIC effects have an important function in enzyme catalysis; many enzyme reactions proceed via charged transition states and so stabilization of charges is a major catalytic factor. The pH dependence of enzyme catalysis depends on the ionization of charged groups sensitive to the surface charge of the protein, and electrostatic interactions involved in hydrogen bonds and salt bridges are important in maintaining the structure of proteins. Gaining a quantitative knowledge of electrostatic interactions is difficult, however, because theoretical analysis based on classical macroscopic electrostatics does not easily adapt to the heterogeneous structure of proteins, and experimental data are rare¹⁻³.

The effect of surface charge on the pK_a values of ionizable side-chains in proteins has been subjected to most analysis, stretching back to the pioneering studies by Linderstrøm-Lang⁴ and Kirkwood and colleagues⁵. A variety of theoretical approaches have been applied in recent years, but there are inconsistencies over basic issues such as whether the protein has a high or low dielectric constant⁶. Experimental evidence is required to test theory and provide a rational basis for designing proteins. Earlier experiments on chemical modifications indicated relatively small effects of surface charge on pK_a values of groups in active sites. For example, acetylation of all surface lysine residues of trypsin raises the pK_a of the active-site histidine by only 0.2 units⁷ and the succinylation of 14 surface lysine residues of chymotrypsin, giving a change in charge of 28 units, raises the pK_a of the active site by only 1 pH unit⁸. In contrast to the earlier studies, we have shown that removal by site-directed mutagenesis of just one surface carboxylate in subtilisin (Asp 99) lowers the pK_a of its active-site histidine (His 64) by

0.3 units at ionic strength 0.1 M, despite the two groups being ~1.3 nm apart⁹. That study, as well as presenting a general method of tailoring the pH-dependence of enzyme activity, indicated that it was possible to measure coulombic interactions individually by a combination of protein engineering and kinetics. We now present an extensive survey based on several mutations, both single and double, that give experimental evidence on the magnitudes of electrostatic effects and how they can be used to tailor the properties of enzymes. The results on altering long range electrostatic interactions show: (1) the high electrostatic shielding of surface charges; (2) the importance of counterion binding to charged groups on the protein and how this has obscured earlier studies; (3), the size of electrostatic effects on changes in pH-activity profiles; (4), the effects on specificity towards charged substrates; (5), the effects on catalysis.

Subtilisin as model system

The serine protease subtilisin (from *Bacillus amyloliquefaciens*) is a good model system for protein engineering studies: its crystal structure, and that of the highly homologous subtilisin from *Bacillus licheniformis*, has been solved at high resolution, as have their complexes with polypeptide inhibitors¹⁰⁻¹²; it has been cloned and expressed in large quantities from plasmids in *Bacillus subtilis*^{9,13}; it may be readily assayed by steady-state kinetics, and absolute rate constants may be obtained after active-site titration. Residue His 64 acts as a general base in catalysis with a pK_a of ~7 (ref. 14). At low pH, His 64 is protonated and so the enzyme is inactive.

We have demonstrated that the pK_a of His 64 can be deter-

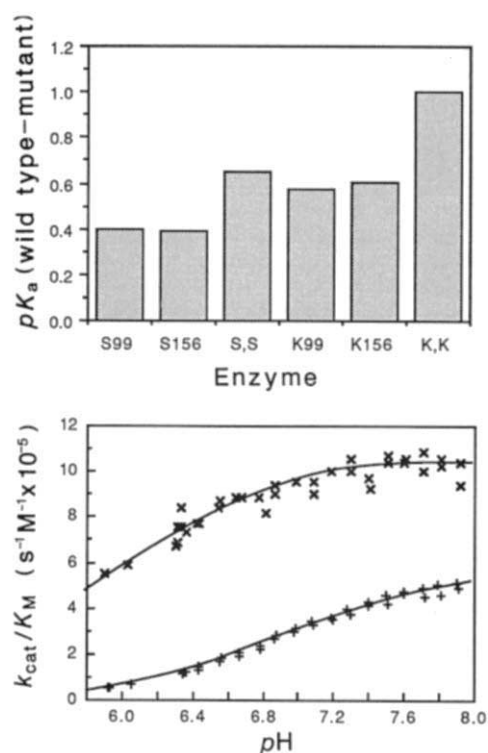


Fig. 1 Top; Shifts in pK_a of His 64 in mutant enzymes relative to wild type at ionic strength 0.001 M (imidazole-imidazolium hydrochloride buffer) and 25 °C. The pK_a of His 64 in wild-type enzyme is 6.91 ± 0.01 . Bottom: pH dependence of k_{cat}/K_M for hydrolysis of succinyl-Ala-Ala-Pro-Phe *p*-nitroanilide by wild-type (+) and double-mutant Asp → Lys 99; Glu → Lys 156 (x) under the same conditions. Mutagenesis was performed as described previously^{9,15}. The mutagenic primers used were as follows: Asp → Ser 99, 5' CGGAACCGCTAGCACCG 3'; Asp → Lys 99, 5' CCGG-AACCTTTAGCACCGA 3'; Glu → Ser 156, 5' GAAGTGCCTGA-GTTACCGG 3'; Glu → Lys 156, 5' AGTGCCTTTGTTACCGG 3'. All mutations were confirmed by DNA sequencing after plaque purification¹⁵. The mutant proteins were expressed in *Bacillus subtilis* DB104 after recloning into pUB110. Purification and kinetics were also performed as described previously¹⁵. Imidazole buffers provide better buffering at low ionic strength and span a better pH range than the phosphate buffers used previously¹⁵. The derived pK_a s consequently have lower standard errors, typically ± 0.01 – 0.02 . The error is higher, however, for the very large shift in the double-lysine mutant (± 0.06).

mined with high precision (typically ± 0.01 – 0.02 pH units) from kinetics (from the pH-dependence of k_{cat}/K_M) so that coulombic interactions may be measured from small shifts in pK_a ¹⁵. The ionization follows a fine titration curve between pH 6 and 8. The data fit precisely a Hill plot of slope 1.00 ± 0.02 under a variety of conditions of ionic strength showing that only a single ionization is being monitored. Two mutants have been made which alter charge in different regions of the protein. One, Glu → Ser 156 removes a negative charge ~ 1.4 – 1.5 nm from the imidazole of His 64—the environment between the two is predominantly aqueous. The other, Asp → Ser 99, removes a charge which is separated from His 64 by 1.2–1.3 nm of an environment which is predominantly protein. The conceptual basis of the experiments is that increasing the overall negative charge on the enzyme should raise the pK_a of the active site histidine by stabilizing the protonated form of His 64 whereas increasing the positive charge should lower the pK_a by destabilizing the protonated form of His 64. The pK_a of His 64 in the mutants is lowered by about 0.4 units at infinitesimal ionic strength¹⁵. Increased

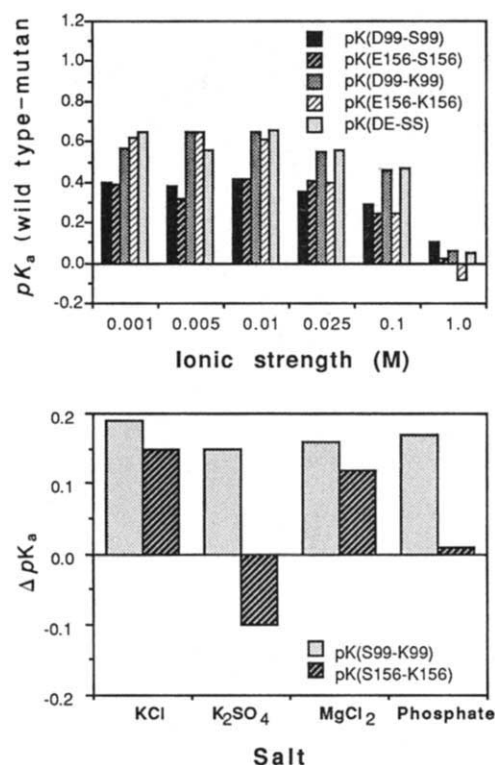


Fig. 2 Effect of ionic strength and nature of buffer on pK_a shift. Top, effect of increasing phosphate buffer from ionic strength 0.005 to 0.1 M followed by addition of KCl to 1.0 M. Bottom, effect of nature of counterion on the pK_a shifts of lysine compared to serine mutants. Ionic strength was maintained at 0.1 M by the electrolyte.

changes in charge have now been made by mutating Asp 99 and Glu 156 to lysine residues. The double mutants Lys 99; Lys 156 and Ser 99; Ser 156 were also constructed.

Electrostatic effects cumulative

Electrostatic interactions are masked by concentrated solutions of electrolytes and so electrostatic effects are most apparent at low ionic strength. Plotted in Fig. 1 are changes in pK_a of His 64 for mutants assayed at an ionic strength of 0.001 M, which is effectively zero (experiments in other buffers give the same shifts in pK_a on extrapolation to zero ionic strength). Changing either Asp 99 or Glu 156 to serine lowers the pK_a by about 0.4 units. Changing both simultaneously to give the double mutant with a change of two charge units lowers the pK_a by 0.65 units. Alternatively, making a double charge change by mutating either Asp 99 or Glu 156 to a lysine lowers the pK_a of His 64 by 0.6 units. The quadruple charge change in the double mutant Asp → Lys 99; Glu → Lys 156 gives a shift of 1.0 units, which is so large that it is leaving the region where quantitative analysis of our data is meaningful. The changes in coulombic interactions are thus cumulative. They are approximately additive for the single mutants, allowing for the long span of the side chain of lysine (~ 0.6 nm). This distance is comparable to the separation of charges in the active site and so there is considerable uncertainty in the distance between the charges: the positively charged ϵ -NH₃⁺ groups of the lysines will tend to stretch away from any positive charge on His 64 and so attenuate the electrostatic effects, especially at low ionic strengths. Indeed, calculations of electrostatic interactions based on the full-extended conformation of lysine suggest that the pK_a -shifts on mutation of Asp or Glu to Ser to Lys are fairly close to being additive (F. Hayes, M. J. E. Sternberg, P. G. Thomas, A. J. R. and A. R. F.,

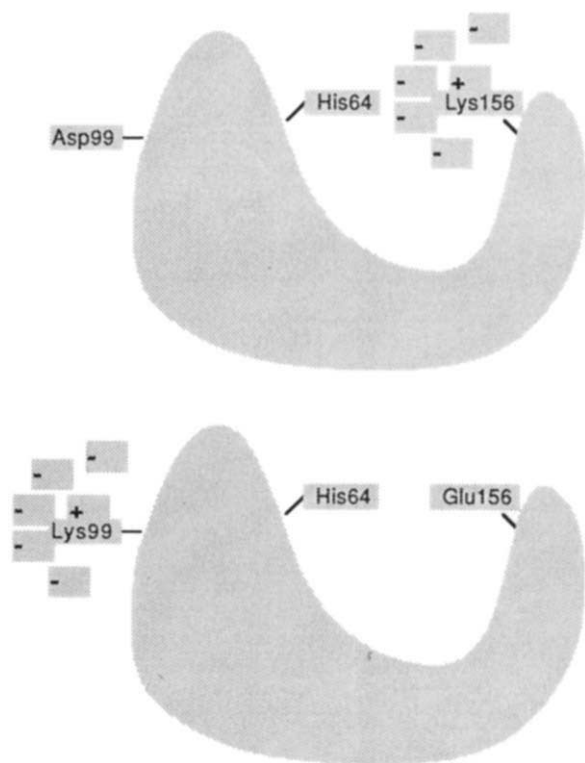


Fig. 3 Schematic representation of ionic atmosphere set up by Lys 99 and Lys 156.

unpublished). The changes for the double mutants appear to be less than additive. Additivity is not necessarily expected, however, because of possible changes in microscopic environment on mutation and changes in interactions between charges mediated by orientation of solvent molecules¹⁵.

High effective dielectric constant

The limiting value of ΔpK_a at low ionic strength may be inserted into the standard electrostatic equations to calculate the effective dielectric constant, D , between the group concerned and His 64 ($\Delta pK_a = 244/Dr$, where r is the separation between two unit charges in Å)¹⁵. In all cases the value is high, independent of whether there is protein or water between the two groups: Asp 99-His 64, $D = 50$; Lys 99-His 64, $D = 80$ –100; Glu 156-His 64, $D = 40$; Lys 156-His 64, $D = 60$ –80. This is not inconsistent with classical macroscopic electrostatics—the dielectric may be replaced by a surface polarization charge which partly neutralizes the charge it surrounds. Warshel and colleagues have also emphasized this point³. They state that, in contrast to common belief, most of the dielectric effect comes from water around the protein and not from the interior of the protein.

Counterion binding

Basic electrostatic theories, such as the classical Debye-Hückel theory, predict that the charge-charge interactions are masked in a general manner by the field set up by counterions in ionic solutions. Examination of Fig. 2 (top) reveals that the difference in pK_a of His 64 between wild-type and mutants Asp $\rightarrow$ Ser 99 and Glu $\rightarrow$ Ser 156 does decrease as the ionic strength of a phosphate buffer is increased to 0.1 M, and finally tends to zero when KCl is added to 1.0 M. But careful examination of the figure reveals that there are specific effects superimposed upon the general. The value of ΔpK_a for the mutant Glu $\rightarrow$ Lys 156 decreases much more rapidly with increasing phosphate concentration than do the other values. We speculated that this results from neutralization of the positive charge on Lys 156 by the binding of phosphate dianion at higher ionic strengths.

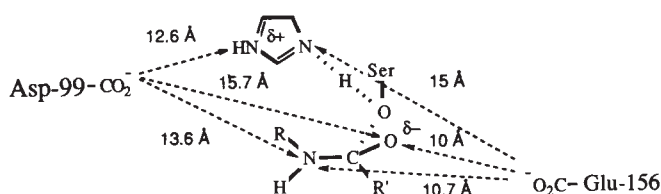


Fig. 4 Stereochemical relationship between charges on Asp 99 and Glu 156 with charges developing on the transition state for the attack of Ser 221 on a peptide substrate. (Crystallographic coordinates kindly provided by Dr M. N. G. James¹²).

Support for this hypothesis comes from the shielding effects of other counterions. The values of pK_a were determined in imidazole-imidazolium hydrochloride buffer at ionic strength 0.001 M, maintaining ionic strength at 0.1 M with electrolytes of different charge types: K^+Cl^- ; $(K^+)_2SO_4^{2-}$; $Mg^{2+}(Cl^-)_2$. The effects are clearly seen when comparing lysine mutants with serine mutants (Fig. 2, bottom). The difference in pK_a between mutants Ser 99 and Lys 99 is roughly independent of electrolyte. Although there is a difference in pK_a of 0.12–0.15 units between mutants Ser 156 and Lys 156 in the presence of either KCl or $MgCl_2$, in phosphate, which is a mixture of phosphate mono- and di-anions, both mutants at position 156 have the same pK_a . Further, in the presence of the full doubly-charged SO_4^{2-} counterion, there is a reversal of the effect—the Lys 156 mutant has a value of pK_a which is 0.1 units higher than that of Ser 156.

The specific effect of negatively-charged ions is clear on consideration of Fig. 3. The ionic atmosphere surrounding a charge on the surface of a protein is asymmetrical, being in the solvent and away from the protein. Thus, the negative ionic atmosphere surrounding a lysine at position 99 is far removed from His 64 because there is protein between Lys 99 and His 64. But, the negative counterions surrounding Lys 156 are concentrated between Lys 156 and His 64 because there is water between them and protein is on the far side. The positive charge on Lys 156 thus tends to concentrate negative ions in the active site. This is far more noticeable for dianions than monoanions because there is twice the coulombic interaction energy between Lys 156 and a dianion for binding, and the dianion when bound exerts twice as large a coulombic interaction energy with His 64. This observation with Lys 156 is not an isolated result; the double lysine mutant is even more sensitive to salts. Mutation of Lys $\rightarrow$ Thr 213 gives also a reversal of expected pK_a change in the presence of moderate concentrations of phosphate, and there is a channel of water between residues 213 and 64 (P. Thomas and A.R.F., unpublished—the effective dielectric in this channel is about 150).

The effects of counterions may explain why the earlier experiments using protein chemistry to modify the lysine residues of chymotrypsin and trypsin gave such low changes in pK_a . The kinetic studies were performed in phosphate buffer of ionic strength 0.1 M or greater, and so the very large changes in surface charge would lead to counterions binding to the active site of the enzymes.

Electrostatic effects and kinetic constants

A simple way of analysing and qualitatively predicting electrostatic effects on kinetics is to ascribe charge distributions to atoms in the transition state and calculate their interaction energies with external charges (see ref. 16). The rate-determining step in the hydrolysis of amides by serine proteases is the formation of an acyl-enzyme with the active site serine¹⁶. This involves the formation of a tetrahedral intermediate in which an oxyanion is formed. The likely regions where charge will be generated in the transition state are on His 64 of the protein as it acts as a general acid-base catalyst, on the carbonyl oxygen of the substrate as it is partly an oxyanion in the transition state,

and, depending whether or not there is rate-determining formation or breakdown of the tetrahedral intermediate, on the N atom of the leaving group¹⁷. Figure 4 illustrates the charges on the transition state for rate-determining formation of the tetrahedral intermediate and the distances of the charges from Asp 99 and Glu 156 (calculated from the crystal structure of the complex between subtilisin and the barley CI-2 inhibitor¹²). We know from our measurements above that the interaction energy of the negative charge on either Asp 99 or Glu 156 with a full positive charge on His 64 is 0.5–0.6 kcal mol⁻¹. We thus expect that the interaction energies with partial charges in the transition state will be somewhat less than this, unless substrate binding lowers the effective dielectric constant. It is seen in Fig. 4 that Asp 99 is closer to the positive charge on His 64 than the negative charge on the oxyanion. Mutation of Asp 99→Ser→Lys should thus destabilize His 64 more than it stabilizes the oxyanion if there is the same effective dielectric constant for the two interactions. That is, the value of k_{cat} should decrease. This is indeed observed: k_{cat} for the hydrolysis of the substrate succinyl-Ala-Ala-Pro-Phe *p*-nitroanilide (suAAPF *p*NA) changes from 57 to 45 to 30 s⁻¹ for these mutations, equivalent to changes in activation energy of 0.14 and 0.24 kcal mol⁻¹ respectively. Glu 156, on the other hand, is closer to the oxyanion. Conversely, the value of k_{cat} should increase on mutation of Glu→Ser→Lys. The value of k_{cat} changes from 57 to 55 to 79 s⁻¹ after these mutations. There is not the monotonic increase found with mutations of Asp 99, but there is a significant and real increase in catalytic rate constant.

There is a straightforward reason why mutations at residue 99 are well-behaved and those at 156 are not. Residue 99 is not in the substrate-binding site and does not interact sterically with the substrate. Glu 156 is in the site and its side chain is close to the side chain of the substrate (Phe) in the primary site. Thus, electrostatic changes of k_{cat} and K_{M} on mutation of 156 will be perturbed by direct steric interactions. It is important to note that the steric effects do not perturb the measurement of pK_{a} by the method employed in this study, the pH dependence of $k_{\text{cat}}/K_{\text{M}}$, as this is purely a function of the free enzyme¹⁶.

Effects on K_{M} can also be qualitatively predicted. The negative charge on the succinyl group of the substrate should interact more favourably with the enzyme as either or both residues are changed to Ser and Lys. On mutation of Asp 99→Ser→Lys, there is a monotonic decrease in K_{M} of 0.17 to 0.13 to 0.10 mM. Mutation of Glu 156→Ser→Lys is more complicated because it is expected that the long side chain of Lys 156 will interact directly, and unfavourably, with the primary side chain of the substrate (Phe). The K_{M} changes from 0.17 to 0.05 to 0.13 mM on mutation. The double mutant Asp→Ser 99; Glu→Ser 156 has a low value of 0.03 mM for K_{M} .

Long range interactions and specificity

Wells *et al.*¹⁷ have shown that the specificity of subtilisin towards charged substrates may be changed where intimate ion pairs are involved. Our data on the effects of mutation on the specificity constant $k_{\text{cat}}/K_{\text{M}}$ towards the negatively-charged substrate suAAPF *p*NA and the positively-charged benzoyl-Val-Gly-Arg *p*-nitroanilide (bz VGR *p*NA) show that long-range electrostatic effects also mediate specificity (Fig. 5). The specificity of the wild-type enzyme for negatively-charged substrate compared to positively-charged is 1.2×10^2 and this increases to 1.6×10^4 for the double mutant Asp→Ser 99; Glu→Ser 156 (ratios of values of $k_{\text{cat}}/K_{\text{M}}$). The change is in the direction logically expected from the effects of charges on the enzyme on charges on the substrate.

More active mutant enzymes

We have generated enzymes that are far more active than wild-type under certain conditions. Although the enhanced activities are most apparent at very low ionic strengths where long range electrostatic effects are most significant, low ionic strength is

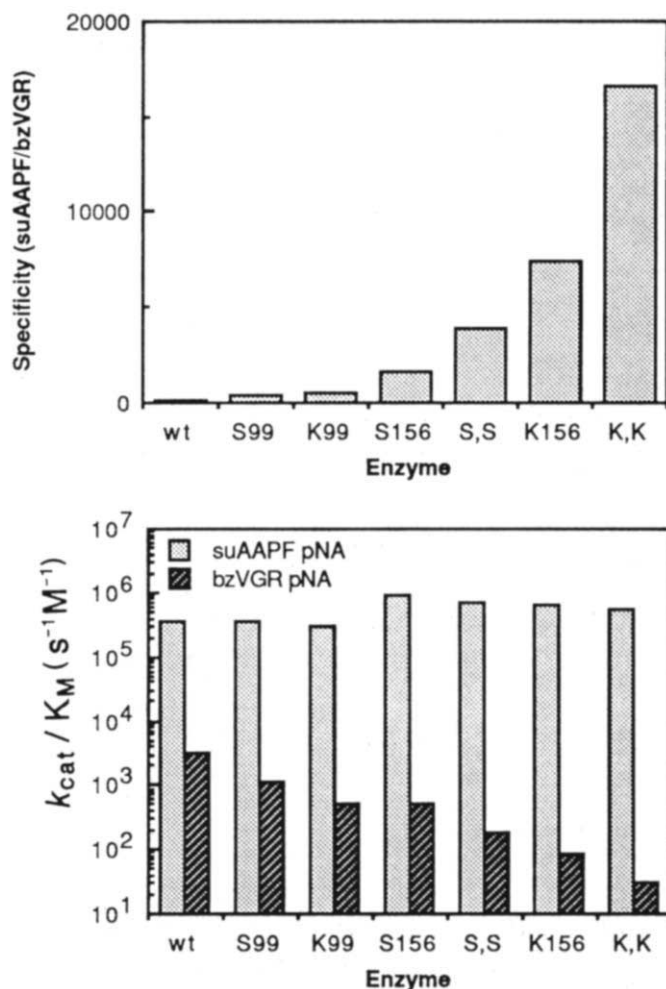


Fig. 5 Specificity of wild-type and mutant enzymes for the substrates succinyl-Ala-Ala-Pro-Phe *p*-nitroanilide and benzoyl-Val-Gly-Arg *p*-nitroanilide at ionic strength 0.01 M (imidazole-imidazolium buffer, pH 7.0) at 25 °C. Top: ratio of values of $k_{\text{cat}}/K_{\text{M}}$ which gives specificity¹⁶. Bottom: absolute values of $k_{\text{cat}}/K_{\text{M}}$.

not uncommon. For example, the aqueous component of a washing machine load, which is the destiny of much of the subtilisin that is produced, has an ionic strength close to zero. The activity of the double mutant Asp→Lys 99; Glu→Lys 156 with suAAPF *p*Na is twice that of wild-type subtilisin at high pH and ten times greater at pH 6 at low ionic strength under $k_{\text{cat}}/K_{\text{M}}$ conditions (Fig. 1).

Rules for changing pH-activity profiles

Our experiments show that modest, but significant, changes in activity may be made by one or two point mutations. By designing a series of mutations based on this knowledge, larger increases may be engineered. But we have found limits on this procedure because of the effects of counterions in solution. The following rules should be applied for designing changes in pH-activity profiles by protein electrical engineering.

(1) Making the surface more negatively charged raises the pK_{a} values of acidic groups in proteins because they all lose a proton on ionization. Conversely, making the surface more positively charged lowers the pK_{a} of all acidic groups.

(2) Changes will be maximized at low ionic strengths.

(3) Significant changes will be manifested at ionic strengths as high as 0.1 M if multiply-charged counterions are avoided.

(4) Mutations should be designed so that they do not concentrate counterions in the active site cavity.

The first two rules are clearly implicit in the classical Linderström-Lang theory. This study has shown that, contrary to previous experimental data, significant effects may be observed at moderate ionic strengths providing rules (3) and (4) are observed. Changes in rate and binding constants may also be engineered by placing charged groups in the proximity of charges in the transition state or substrate. As most enzymes use general acid-base catalysis and their reactions involve

charged transition states, the rational modification of catalytic properties by engineering electrostatics should be generally applicable.

This work was funded by the SERC (Protein Engineering Initiative). A.J.R. has a studentship from the Biotechnology Directorate of the SERC. We thank Dr M. N. G. James for unpublished coordinates of the subtilisin-CI-2 complex.

Received 5 May; accepted 29 June 1987.

1. Rogers, N. K. *Prog. Biophys. molec. Biol.* **48**, 37-66 (1986).
2. Matthew, J. B. A. *Rev. Biophys. Bioengng* **14**, 387-417 (1985).
3. Warshel, A. & Russell, S. T. *Q. Rev. Biophys.* **17**, 283-422 (1984).
4. Linderström-Lang, K. *C. r. Trav. Lab. Carlsberg* **15**, 70-76 (1924).
5. Tanford, C. & Kirkwood, J. G. *J. Am. Chem. Soc.* **79**, 5333-5339 (1957).
6. Warshel, A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 444-448 (1984).
7. Spomer, W. E. & Wootton, J. F. *Biochim. biophys. Acta* **235**, 164-171 (1971).
8. Valenzuela, P. & Bender, M. L. *Biochim. biophys. Acta* **250**, 538-548 (1971).
9. Thomas, P. G., Russell, A. J. & Fersht, A. R. *Nature* **318**, 375-376 (1985).

10. Wright, C. S., Alden, R. A. & Kraut, J. *Nature* **221**, 235-242 (1969).
11. Drenth, J. & Hol, W. G. J. *J. molec. Biol.* **28**, 543-551 (1967).
12. McPhalen, C. A. & James, M. N. G. *Biochemistry* **26**, 261-269 (1987).
13. Wells, J. A., Ferrari, E., Henner, D. J., Estell, D. A. & Chen, E. Y. *Nucleic Acids Res.* **11**, 7911-7925 (1983).
14. Phillip, M., Tsai, I.-H., & Bender, M. L. *Biochemistry* **18**, 3769-3773 (1979).
15. Russell, A. J., Thomas, P. G. & Fersht, A. R. *J. molec. Biol.* **193**, 803-813 (1987).
16. Fersht, A. R. in *Enzyme Structure and Mechanism* (Freeman, New York, 1985).
17. Fastrez, J. & Fersht, A. R. *Biochemistry* **12**, 1067-1074 (1973).
18. Wells, J. A., Powers, D. B., Bott, R. R., Graycar, T. P. & Estell, D. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1219-1223 (1987).

LETTERS TO NATURE

A direct determination of linear-size evolution of elliptical radio galaxies

M. J. A. Oort*, P. Katgert* & R. A. Windhorst†

* Sterrewacht Leiden, PO Box 9513, 2300 RA Leiden, The Netherlands

† Mount Wilson & Las Campanas Observatories, 813 Santa Barbara Street, Pasadena, California 91101, USA

A redshift dependence of the maximum distance out to which active galactic nuclei (AGN) can transport their relativistic particles was first found for radio quasars by Miley¹. He demonstrated that this distance was, on average, smaller in the past than it is at present. Analysing the run of median angular size with flux density for the entire extragalactic radio source population, Kapahi² concluded that the radio galaxy population probably shares this property with the extended radio quasars. By using samples of radio ellipticals complete to vastly different flux limits, we can, for the first time, separate redshift- and luminosity-dependences of their linear sizes. Ram pressure in a medium with cosmologically varying density will naturally cause a dependence on redshift³, but this cannot explain so steep a dependence as that implied here. It may be expected if galaxy haloes influence the outward energy transport⁴.

The variation of the linear sizes of radio sources with redshift provides us with important information on the interaction between the outward flow of energy and the surrounding medium, and therefore on the epoch dependence of the properties of the medium. Direct determinations of the effect, like those based on extended radio quasars^{1,5}, may be influenced by a dependence of linear sizes on radio luminosity through luminosity-redshift correlations in flux-limited samples. Indirect determinations, based on the median angular (as opposed to linear) size of the entire radio-source population and especially on its variation with flux density^{2,6}, are generally taken to imply a rather mild redshift dependence of linear sizes of the entire source population. But this conclusion (and, in particular, the quantitative details) depends on the assumptions that have to be made about the spatial distribution of the sources that define the median angular size at each flux density. Attempts at determining the redshift dependence of linear size of the entire source population have been made as well^{13,14}. These studies find only a mild dependence of size on redshift. But they may be biased because the samples are flux limited or have rather incomplete redshift information.

Here we concentrate on the sizes of elliptical radio galaxies, since these galaxies form a homogeneous class that is well suited for a direct determination of systematic variations of linear sizes with redshift and radio luminosity. For our purpose, the prime virtue of elliptical radio galaxies in the present context is that they show a very small dispersion in absolute magnitude, and that their mean absolute magnitude varies very gradually with radio luminosity⁷. Therefore, we can estimate quite reliable redshifts from apparent magnitudes for those elliptical radio galaxies for which no spectroscopic redshift is available⁸.

We have recently obtained high resolution Very Large Array (VLA) images of 35 radio ellipticals with a median flux density of ~5 mJy at 1.4 GHz, as a part of a programme to study the morphology of faint radio sources⁹. One third of these ellipticals have spectroscopic redshifts and, for the remainder, redshifts were estimated from high-quality photometry⁸. Almost all of them have radio luminosities (P) in the range of $\log P[\text{WHz}^{-1}]$ at 1.4 GHz from 24 to 26. Only three of these elliptical radio galaxies remain unresolved at the highest VLA resolution of 20-cm wavelength (<1 arc s).

Comparable data were obtained by Machalski and Condon¹⁰ for a sample of 46 elliptical radio galaxies with a median flux density of ~400 mJy. In their case, only five spectroscopic redshifts are available. Because their magnitude estimates are less accurate than for the sources in our sample, we have only used their ellipticals with estimated redshifts less than 0.4, as opposed to the maximum value of ~0.75 in our sample.

In addition to the 81 elliptical radio galaxies with $\log P$ between 24 and 27 and redshifts between ~0.1 and ~0.75 from these two samples, we have used the 82 elliptical radio galaxies in the same radio luminosity range, but all with redshifts less than 0.1, discussed by Gavazzi and Perola¹¹. All galaxies in this last sample have spectroscopic redshifts. Note that all three samples were observed with sufficient resolution to resolve nearly all of them. Furthermore, care was taken not to underestimate the largest angular size due to resolving low surface-brightness components at high resolutions. If a substantial difference was seen between low- and high-resolution maps of a source, the maximum angular size was always taken.

Linear sizes were determined for the 163 elliptical radio galaxies using the method described by Oort *et al.*⁹. For extended sources, which generally show brightening towards the outermost edges, the (largest) distance between these outermost components was taken. For resolved (not extended) sources that generally consist of a single diffuse component, the full width at half maximum was used. Some published values^{10,11} for diffuse and not very collimated sources had to be revised to ensure the maximum consistency of the size determinations.

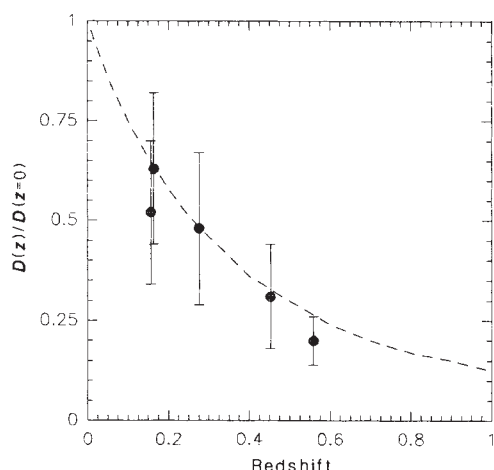


Fig. 1 The median linear size of elliptical radio galaxies as a function of redshift. Sizes have been normalized to a local ($z=0$) median value at the same radio luminosity. The dotted line represents the best fit: $D_{\text{med}}(z) \propto (1+z)^{-3.3}$.

Using this method a complication may occur when the morphological type of radio sources at a given radio luminosity changes with redshift. Locally, there is a critical radio luminosity $\log P = 24.5$ above which sources are primarily doubles (extended) and below which a significant fraction of the elliptical radio galaxies is resolved. If this critical radio luminosity is larger at higher redshifts, one may introduce an artificial z -dependence, caused by a changing morphological content rather than a general decrease in size of one class of objects. However, no such change in morphology is seen in our data. Above $\log P = 24.5$, 80–90% of the elliptical radio galaxies are doubles, at all redshifts sampled by us. We can therefore safely use the method sketched here.

In each of the three decades of radio luminosity ($24 \leq \log(P[\text{W Hz}^{-1}]) \leq 27$), we have calculated the ratio between the median linear size in the redshift interval centred on specific redshifts and the local ($z=0$) median linear size in the Gavazzi and Perola sample. This procedure completely separates redshift and luminosity dependences of linear size. The values thus obtained are displayed in Fig. 1. To first order, the elliptical galaxies of all radio luminosities sampled consistently define a single redshift dependence which we have (arbitrarily) represented by $D_{\text{med}}(z) \propto (1+z)^{-n}$. We then find $n = 3.3 \pm 0.5$. A fit to a function $e^{m\tau}$ (with τ = look-back time), is slightly less satisfactory. Note that our choice of q_0 assures that we have minimized the strength of the redshift-dependent decrease. The point at the highest redshift lies somewhat below the fit to the data. This may be caused by a possible underestimation of some of the angular sizes in this redshift bin by about 20%, because there are slightly more sources above $z=0.5$ which are resolved.

If we now use this parameterization, we can derive the single $\log P$ -dependence defined by the elliptical radio galaxies at all redshifts sampled. By reducing all D values to a single redshift, we find:

$$D_{\text{med}}(P) \propto P^{0.3+0.05}$$

Before comparing our result with theoretical predictions, we want to stress the fact that our fit of $D(z, \log P)$ describes only the state of an entire population at a certain epoch, but does not reflect the physical evolution of individual objects with time. A physical interpretation of our result requires analysis of both $D(t)$ and $\log P(t)$.

In 1971, de Young³ predicted that the maximum distance to which the radio source produced by an AGN can expand should vary approximately as $(1+z)^{-0.8}$, for a given luminosity and with ram pressure exerted by a medium with density propor-

tional to $(1+z)^3$. Others have, in a qualitative way only, invoked Compton 'snuffing' which in the standard cosmological model increases strongly with redshift. Recently, Gopal-Krishna and Wiita⁴ have constructed a model in which the pressure balance between the galaxy halo and the intergalactic medium defines a transition surface, the position of which is redshift-dependent. From this they predict a maximum distance up to which the jet remains collimated that varies approximately as $(1+z)^{-3}$ up to redshifts, $z \sim 1$. This is in excellent agreement with our empirically determined relation.

Finally, one might wonder whether our result is not inconsistent with Kapahi's models, which represent the observed median angular sizes quite satisfactorily over many decades of flux density on the basis of a considerably milder redshift dependence than we find ($\sim (1+z)^{-1.5}$). But as will be discussed in detail elsewhere¹², this apparent inconsistency disappears with the newest results on the median angular size below ~ 50 mJy at 1.4 GHz. From these it appears that, until now, the median angular size of the sources < 50 mJy has been systematically (and substantially for the fainter sources) overestimated by all observers as a result of insufficient resolution.

M.J.A.O. thanks ASTRON/ZWO for financial support.

Received 16 March; accepted 28 May 1987.

1. Miley, G. K. *Mon. Not. R. astr. Soc.* **152**, 477–490 (1971).
2. Kapahi, V. K. *Mon. Not. R. astr. Soc.* **172**, 513–533 (1975).
3. de Young, D. S. *Astrophys. J.* **167**, 541–551 (1971).
4. Gopal-Krishna & Wiita, P. *J. Mon. Not. R. astr. Soc.* (in the press).
5. Hooley, A., Longair, M. S. & Riley, J. M. *Mon. Not. R. astr. Soc.* **182**, 127–145 (1978).
6. Kapahi, V. K. *Highlights of Astronomy* (ed. Swings, J. P.) 371–376 (Reidel, Dordrecht, 1986).
7. Auremma, C. et al. *Astr. Astrophys.* **57**, 41–50 (1977).
8. Kron, R. G., Koo, D. C. & Windhorst, R. A. *Astr. Astrophys.* **146**, 38–58 (1985).
9. Oort, M. J. A., Katgert, P., Steeman, F. W. M. & Windhorst, R. A. *Astr. Astrophys.* **179**, 41 (1987).
10. Machalski, J. & Condon, J. J. *Astr. J.* **90**, 973–986 (1985).
11. Gavazzi, G. & Perola, G. *G. Astr. Astrophys.* **66**, 407–416 (1978).
12. Oort, M. J. A. *thesis*, Univ. Leiden (1987).
13. Katgert-Merkelijn, J., Lari, C. & Padrielli, L. *Astr. Astrophys. Suppl.* **40**, 91–106 (1980).
14. Eales, S. A. *Mon. Not. R. astr. Soc.* **217**, 179–203 (1985).

Decay of primordial cosmic rotation in inflationary cosmologies

Öyvind Grøn & Harald H. Soleng*

Oslo College of Engineering, Cort Adelers Gate 30, Oslo, Norway
and Institute of Physics, University of Oslo, Oslo, Norway

* Institute of Theoretical Astrophysics, University of Oslo, Oslo, Norway
and Institute of Physics, University of Oslo, Oslo, Norway

Observations have shown that within very strict limits the plane of oscillation of a Foucault pendulum does not rotate with respect to the distant galaxies. We may take the plane of the pendulum as a definition of non-rotation. In this sense our Universe is rotating very slowly, if at all. Ernst Mach advocated that this was no coincidence¹, and the vanishing cosmic rotation is part of the empirical basis of what is known as Mach's principle. Amongst other things it decrees that cosmic rotation should be exactly zero. Assuming that the Universe came to being as a mini-universe of Planck dimensions which went directly into an inflationary epoch driven by a scalar field with a flat potential, we argue that the cosmic vorticity has decayed by a factor of about 10^{-145} , due to the non-rotation of the false vacuum and the exponential expansion during inflation. Here we show that the hypothesis that the rotation of galaxies is caused by cosmic vorticity is not compatible with the inflationary model.

The rotation of planets, stars and galaxies made Gamow² speculate that the Universe is rotating and that the observed

rotation of stars and galaxies could be a result of cosmic vorticity. Later Gödel³ gave his famous example of a rotating cosmological solution of the field equations of Einstein. He further discussed the possibility of a cosmic explanation of galactic rotation. This hypothesis we will call the Gamow/Gödel hypothesis.

The Gödel solution settled the question of whether Mach's principle is fully incorporated in the general theory of relativity. The assumption that it is fulfilled in finite models was also proved wrong by an explicit finite rotating universe model⁴.

Thus we cannot *a priori* exclude rotating solutions. Accepting the possibility of cosmic rotation, one would like to have an explanation of the slow rotation of the present Universe.

The effect of vorticity on the microwave background was analysed by Collins and Hawking⁵ and by Barrow *et al.*⁶ who found the following limit on the period of rotation, $T_{\text{obs}} > 3 \times 10^5 T_H$ (where $T_H = 1.2 \times 10^{10}$ years is the Hubble time) in the case $k = 0$. The limit is even higher if space is closed.

Ellis and Olive⁷ pointed out that inflation could solve the problem of why the Universe rotates so slowly. They considered the effect of the exponential expansion on vorticity. This approach has, however, been criticized because Ellis and Olive have neglected the effect of shear on the evolution of vorticity⁸.

In the literature there are two solutions which are presented as rotating generalizations of de Sitter space (Vaidya⁹, Grön¹⁰). The first of these vacuum solutions does not allow twisting geodesic motion of the fundamental observers. The geodesic condition would lead to $m = 0$, which, in the Vaidya solution, corresponds to $\omega = 0$; vanishing vorticity. The solution of Grön has been criticized by Jantzen¹¹, and cannot be said to be rotating.

The discussion here will not be based on any specific universe model, and it will be valid to all inflationary models driven by a homogeneous scalar field with a flat potential, like the chaotic inflationary universe of Linde¹².

In a complete cosmological theory¹³ the initial conditions will have to be explained within the theory. The simplest example of such a theory is the idea of the Universe as a quantum fluctuation in which all conserved quantities have a net value equal to zero¹⁴. The most successful attempts in the direction of a complete model is the theory of the wavefunction of the Universe^{15,16}. Considering shear anisotropy the wavefunction is strongly peaked around isotropy¹⁷⁻¹⁹. Furthermore, a recent mini-superspace investigation²⁰ has shown that rotation is exponentially suppressed.

Suppression of initial cosmic rotation probably holds generally. Here we therefore assume an initial vorticity not greater than the Planck vorticity discussed by Panov²¹.

Assuming a barotropic equation of state of the cosmic perfect fluid Sanz²² has found the following vorticity evolution equation:

$$\omega^2 \{ \ln [(\rho + P)\omega R^5] \}' = \sigma_{\alpha\beta} \omega^\alpha \omega^\beta \quad (1)$$

where $\sigma_{\alpha\beta}$, ω^α , ω , and R are the shear tensor, the vorticity vector, the vorticity scalar and the cosmic scale factor respectively. If the component of the shear is zero along the axis of rotation, angular momentum is conserved. Shear is decaying as R^{-3} during expansion. As a result of this, shear can be neglected in the vorticity evolution equation, if the initial shear was not too strong. If the fluid apart from the scalar field has the equation of state $P = \rho/3$, the vorticity will decay as R^{-1} during inflation. At the end of the inflationary epoch the scale factor has increased by a factor of 10^{29} (to solve the flatness problem²³), and the cosmic vorticity has decayed by a factor 10^{-29} . At the same time the energy density of the radiation has decayed by a factor of 10^{-116} .

The scalar field is assumed to be homogeneous as the wavefunction in the mini-superspace models is peaked around homogeneity²⁴. Let φ be a scalar field with a lagrangean

$$\mathcal{L} = \frac{1}{2} \varphi_{,\alpha} \varphi^{,\alpha} - V(\varphi) \quad (2)$$

If the potential is sufficiently flat it will act as a cosmological

constant. Introducing a four-velocity of the scalar field²⁵ proportional to the conjugate momentum we find

$$u_\mu = \varphi_{,\mu} (\varphi_{,\alpha} \varphi^{,\alpha})^{-1/2} \quad (3)$$

$\varphi_{,t} = \varphi_{,t}$ is the only nonvanishing component of $\varphi_{,\mu}$. Hence we get the four-velocity

$$u = \frac{1}{\sqrt{g^{tt}}} dt \quad (4)$$

As the metric will have to be homogeneous equation (4) leads to $du = 0$. The rotation vector is given by²⁶

$$\omega^\mu = (\star du \wedge u)^\mu \quad (5)$$

where $\star$ is the Hodge $\star$ -operator. We may thus conclude that the scalar field does not take part in the rotation. As the false vacuum energy density remains approximately constant during inflation, the non-rotating matter becomes a factor of 10^{116} denser than the rotating fluid, if they were of equal density at the Planck time. At the end of the inflationary era the false vacuum turns into radiation. The density, pressure, and vorticity of this 'new' radiation we denote by ρ_v , P_v , ω_v , whereas the original radiation is denoted by a subscript r , and the totals are without subscripts. Local conservation of angular momentum during decay of the false vacuum leads to

$$(\rho + P)\omega R^5 = (\rho_v + P_v)\omega_v R^5 + (\rho_r + P_r)\omega_r R^5 \quad (6)$$

As the vorticity of the false vacuum is zero, $\omega_v = 0$. Furthermore, $\rho_v \gg \rho_r$, and $P = \rho/3$. We then find

$$\omega = \frac{\rho_r}{\rho_v} \omega_r \quad (7)$$

which implies that the cosmic vorticity is diluted by an extra factor 10^{-116} , due to the relative density of the rotating fluid compared to the non-rotating decay products of the false vacuum, leading to a total decay of $\sim 10^{-145}$. This diluting effect is by far the most important effect in the decay of rotation.

If shear is important in the vorticity evolution equation, the decay due to expansion may be less effective than $1/R$, but the diluting effect is enough to solve the rotation problem. Thus this effect alone will reduce vorticity by a factor 10^{-116} . Assuming an initial Planck rotation, and the $1/R$ decay due to expansion, the rotation period of the Universe after inflation is $T = 10^{145-43} \text{ s} = 10^{102} \text{ s}$.

During the radiation epoch the Universe has expanded a factor 10^{21} , and finally in the matter-dominated period it has expanded by a factor of 10^4 . In the radiation epoch, vorticity decays as $1/R$, and in the matter period there is a $1/R^2$ dependence. Thus we will expect a present period $T_0 = 10^{102+21+8} \text{ s}$ which adds up to $T_0 = 10^{131} \text{ s} = 10^{124} \text{ yr}$. Even if the uncertainty with respect to the effect of the shear is taken into account, we can conclude that the period of rotation must be more than 10^{95} years. One should notice that the expansion during inflation may very well be greater than 10^{29} , leading to an even greater period of rotation of the Universe.

This cosmic vorticity is far too slow to be observed directly. The Gamow/Gödel hypothesis that galactic rotation is a result of cosmic vorticity is clearly inconsistent with this result, as the typical period of a galaxy is only 10^{14} s . At the scale of the Local Supercluster Aaronson *et al.*²⁷ have found a rotation velocity of $180 \pm 58 \text{ km s}^{-1}$ at the position of the Local Group. The distance to the centre of the Local Supercluster is 19.5 Mpc which gives a rotation period of 10^{19} s . We conclude that the Gamow/Gödel hypothesis can explain rotation neither on the galactic nor on the supergalactic scale.

Received 23 April; accepted 28 May 1987.

1. Mach, E. *Die Mechanik in Ihrer Entwicklung Historisch-Kritisch Dargestellt* (1883); *The Science of Mechanics* 6th edn (Transl. McCormack, T. J.) (Open Court, LaSalle, Illinois 1960).
2. Gamow, G. *Nature* **158**, 549 (1946).
3. Gödel, K. *Rev. mod. Phys.* **21**, 447-450 (1949).

4. Ozsváth, I. & Schücking, E. L. *Ann. Phys.* **55**, 166–204 (1969).
5. Collins, C. B. & Hawking, S. W. *Mon. Not. R. astr. Soc.* **162**, 307–320 (1973).
6. Barrow, J. D., Juszkiewicz, R. & Sonoda, D. H. *Mon. Not. R. astr. Soc.* **213**, 917–943 (1985).
7. Ellis, J. & Olive, K. A. *Nature* **303**, 679–681 (1983).
8. Rothmann, T. & Ellis, G. F. R. *Phys. Lett.* **B180**, 19–24 (1986).
9. Vaidya, P. C. *Pram. J. Phys.* **25**, 513–518 (1985).
10. Grön, Ö. *Phys. Rev. D* **33**, 1204–1205 (1986).
11. Jantzen, R. T. *J. math. Phys.* **27**, 2748–2750 (1986).
12. Linde, A. D. *Phys. Lett.* **129B**, 177–181 (1983).
13. Grishchuk, L. P. & Zeldovich, Ya. B. in *Quantum Structure of Space and Time Proc. Nuffield Workshop Imperial College, London, 3–21 August 1981*, 409–422 (eds Duff, M. J. & Isham, C. J.) (Cambridge University Press, 1982).
14. Tryon, E. P. *Nature* **246**, 396–397 (1973).
15. Hartle, J. B. & Hawking, S. W. *Phys. Rev. D* **28**, 2960–2975 (1983).
16. Fang, L. Z. & Wu, Z. C. *Int. J. mod. Phys. A* **1**, 887–912 (1986).
17. Hawking, S. W. & Luttrell, J. C. *Phys. Lett.* **143B**, 83–86 (1984).
18. Wright, W. A. & Moss, I. G. *Phys. Lett.* **154B**, 115–119 (1985).
19. Amsterdamski, P. *Phys. Rev. D* **31**, 3073–3078 (1985).
20. Fang, L. Z. & Mo, H. J. *Phys. Lett.* **B186**, 297–302 (1987).
21. Panov, V. F. *Soviet Phys. J.* **28**, 981–984 (1985).
22. Sanz, J. L. *J. math. Phys.* **23**, 1732–1736 (1982).
23. Guth, A. H. *Phys. Rev. D* **23**, 347–356 (1981).
24. Halliwell, J. J. & Hawking, S. W. *Phys. Rev. D* **31**, 1777–1791 (1985).
25. Belinskii, V. A. & Khalatnikov, I. M. *Soviet Phys. JETP* **36**, 591–597 (1973).
26. Ryan, M. P. & Shepley, L. C. *Homogeneous Relativistic Cosmologies* (Princeton University Press, 1975).
27. Aaronson, M., Huchra, J., Mould, J., Schechter, P. L. & Tully, R. B. *Astrophys. J.* **258**, 64–76 (1982).

Regularly pulsed neutrinos from supernova SN1987A?

Martin Harwit*, Peter L. Biermann†, Hinrich Meyer‡§ & Ira M. Wasserman*

* Astronomy Department, Cornell University, Ithaca, New York 14853, USA

† Max Planck Institut für Radioastronomie, D-5300 Bonn, FRG

‡ Fachbereich Physik, Universität Wuppertal, D-56 Wuppertal, FRG

The neutrino events from SN1987A observed with the Kamiokande and IMB experiments^{1,2} appear to show evidence of a period (P) of approximately 8.9 milliseconds. While the statistical significance of this period is marginal in each individual experiment, the two sets of data are compatible. Here we remark on a few consequences of interest should the periodicity be confirmed by observations. Interpreting the apparent period $P \approx 8.9$ ms as a rotation of a compact object would imply that: (1) neutrino emission is anisotropic; (2) the neutrino mass, averaged over all flavours observed is less than $0.2 \text{ eV}/c^2$. We also note that $P = 8.9$ ms is a reasonable period for very young pulsars.

The recent detection of a neutrino burst from SN1987A by two independent experiments, Kamiokande and IMB (Hirata *et al.*¹, Bionta *et al.*²) for the first time permits an observationally based discussion of events that take place in the core of an exploding star. Theoretical expectations have been summarized by Woosley and Weaver³; the effects of magnetic fields and rotation have been computed and discussed by LeBlanc and Wilson⁴, Müller and Hillebrandt⁵ and Smarr *et al.*⁶ among others. Meier *et al.*⁷ predict a rotation period for the compact core of ~ 8 ms.

Searching for periods in a sparse set of arrival times t_j ($j = 1, 2, \dots, N$) is a somewhat subjective enterprise. We have chosen to 'fold' the data relative to various candidate periods P_α chosen according to arbitrary rules and then we have computed the mean square residual

$$y_\alpha = \frac{1}{N} \sum_{j=1}^N x_{j\alpha}^2 \quad (1a)$$

where, if $N(Z)$ is the nearest integer to Z

$$x_{j\alpha} = \frac{t_j}{P_\alpha} - N\left(\frac{t_j}{P_\alpha}\right) \quad (1b)$$

For random arrival times each $x_{j\alpha}$ is uniformly distributed on $[-1/2, +1/2]$ and the distribution function for y_α is

$$f_N(y_\alpha) = \int_{-1/2}^{+1/2} dx_{1\alpha} \dots dx_{N\alpha} \delta\left(y - \sum_j x_{j\alpha}^2/N\right) \quad (2)$$

which depends on N but not α . It is easily shown that $\langle t_\alpha \rangle = 1/12$ and $\langle (y_\alpha - 1/12)^2 \rangle = 1/180N$, so that $f_N(y_\alpha)$ becomes sharply peaked as $N \rightarrow \infty$ but is fairly broad for modest N . At small values of y_α , a good approximation to the random distribution is

$$f_N(y_\alpha) \approx \frac{(N\pi)^{N/2} y_\alpha^{N/2-1}}{(N/2-1)!} \quad (3)$$

so that the probability that $y_\alpha < Z$ is

$$P_N(<Z) \approx \frac{(NZ\pi)^{N/2}}{(N/2)!} \quad (4)$$

We can use equations (1)–(4) to evaluate the probability that a single period provides a 'good fit' (defined arbitrarily) to two independent sets of arrival times, respectively, that is, those for the IMB and Kamiokande neutrino events.

In order to be able to do this we must first adopt a specific period search strategy. We begin by testing candidate periods to millisecond accuracy

$$P_\alpha^{\text{IMB}}(\text{ms}) = 6,000/n_\alpha \quad (5)$$

on the IMB relative arrival times kindly provided for us by the IMB collaboration: 0, 412, 650, 1,141, 1,562, 2,684, 5,010 and 5,582 ms. Equation (5) represents a slight oversampling of possible periods. The integer n_α spans the interval $1 \leq n_\alpha \leq 3,000$; it is easily shown that y_α is the same for n_α and $6,000 - n_\alpha$. By our choice of periods, see equation (5), there are seven independent arrival times. We have numerically identified all periods with $n_\alpha \leq 849$ and $y_\alpha < 0.02$, finding three such periods at $n_\alpha = 85, 304$ and 673 , corresponding to $P_\alpha^{\text{IMB}} = 70.59, 19.74$ and 8.915 ms, respectively. Values of $n_\alpha > 849$ were not tested because the implied timing residual would be < 1 ms (the accuracy of the data). Equation (4) gives $P_7(<0.02) \approx 4.85 \times 10^{-3}$, leading to an expected number of good fits, $N_{\text{ex}} \approx (849)(4.85 \times 10^{-3}) = 4.1$ for random arrival times, which is consistent with the numerical results within statistical accuracy. A different number of good fits may, of course, be obtained by either choosing a different 'good fit' threshold or by examining an alternative set of candidate periods.

To identify 'good fit' periods in the Kamiokande data, we test periods defined by

$$P_\alpha^{\text{K}}(\text{ms}) = 135,000/n'_\alpha \quad (6)$$

on Kamiokande arrival times accurate to 0.1 ms kindly provided for us by Dr Y. Totsuka of the Kamiokande consortium: 0, 106.8, 302.4, 323.3, 506.9, 685.2, 1,540.2, 1,727.9, 1,914.5, 9,218.8, 10,432.2 and 12,438.8 ms. There are 67 different values of n'_α that, when scaled to the IMB data, round off to the 3 integers n_α that correspond to the 'good fit' IMB periods. Of these 67 candidate periods we identify those with $y_\alpha < 3.07 \times 10^{-2}$. These fits are equally as good as their IMB counterparts in the sense that $P_{11}(<3.07 \times 10^{-2}) = P_7(<0.02) \approx 4.85 \times 10^{-3}$. Five 'good fits' were identified in this way at $n'_\alpha = 15,141, 15,142, 15,151, 15,152$ and $15,153$, corresponding to $P_\alpha^{\text{K}} = 8.9162, 8.9156, 8.9103, 8.9097$ and 8.9091 ms respectively.

For random arrival times, the expected number of such good fits is $N_{\text{ex}} = (67)(4.85 \times 10^{-3}) = 0.32$, and the probability of finding five good fits at this level is $P_5 = 1.9 \times 10^{-5}$. Perhaps more striking is that all of the good fits occur in a single period bin $P \approx 8.91$ – 8.92 ms, corresponding to a single good fit IMB period. This is a circumstance that would arise about 1.2% of the time randomly given that there are five good-fit periods among the 67 candidates.

We emphasize that the probabilities quoted above refer only

§ Present address: Santa Cruz Institute for Particle Physics, University of California, Santa Cruz, California 95064, USA.

to a specific period search strategy. Any strategy based on residuals must include (1) enumeration of candidate periods, (2) thresholds for acceptance and (3) criteria for accepting periods in independent data sets. It is easy to imagine strategies that would result in relatively frequent period coincidences. Two physically motivated changes in any adopted search strategy that increase the number of good fits to independent data sets may be envisaged. First, for any realistic beam pattern we expect non-zero residuals even if the emission is strictly periodic, so acceptance thresholds may not be set arbitrarily low. Second, period changes over the detection period may occur, further increasing the expected residuals even for periodic emission as well as increasing the set of 'closely matched' periods in independent data sets. One can always 'improve' the fit for a given model by including a period-change parameter $\Delta P/P$ or $\dot{P}$, as there are many more candidate models in the resulting two-dimensional parameter space than in the original one-dimensional space. Fractional period changes $\Delta P/P \leq 0.10$ are expected because of neutrino emission during gravitational collapse⁸.

We worry, of course, that we could be deceived by peculiar instrumental effects either in the detector or in the clock. But because the detector has a dead time of less than 1 μ s and a multi-hit capability, it appears unlikely that the apparent periodicity is due to the detector. A detailed discussion with Dr Y. Totsuka of the Kamiokande team similarly revealed that no other instrumental effect (for example, a clocking effect) is able to explain the apparent periodicity. Bearing in mind the statistical ambiguities that are inherent in identifying periodicities in a sparse set of arrival times, we consider below whether or not periodic neutrino emission with $P \approx 8.9$ ms is physically plausible.

At the centre of a collapsed supernova, magnetic fields and rotation rates may attain extreme values. Electron neutrinos and antineutrinos trapped in the core of the star through elastic scattering off nucleons can create extreme pressures. Convective instabilities may arise, as shown by Smarr *et al.*⁶, and transportation of neutrinos and antineutrinos from the interior of the core to the vicinity of the neutrinosphere may be significantly more efficient through convection than by means of lengthy random walks. Do all these circumstances suggest that a rapidly rotating core might emit evenly spaced bursts of neutrinos in the direction of an observer? Alternatively, should we expect that the flux received will be steady? Or should we think that the received beam will vary chaotically during the brief period of peak neutrino emission?

Periodic emission with even modest beaming is difficult to achieve in models that rely purely on neutrino transport anisotropies due, for example, to strong magnetic fields. A fundamental problem shared by all such models is that neutral current interactions scatter neutrinos and antineutrinos effectively. For all neutrino and antineutrino flavours, the mean free path for coherent scattering in a core of density ρ_{12} , measured in units of 10^{12} g cm⁻³, is only $0.24 \rho_{12}^{-5/3}$ km (Shapiro and Teukolsky⁹). As one approaches nuclear densities $\sim 3 \times 10^{14}$ g cm⁻³ the neutrinos and antineutrinos are unable to escape the core without undergoing a large number of collisions which would destroy any initial emission anisotropies.

Similarly, a mechanism like von Zeipel's gravity-brightening

(Tassoul¹⁰) due to the distortion of a rapidly rotating, strongly magnetized star can be dismissed, because neither rotation with a 8.9 ms period, nor magnetic fields with strengths below 10^{14} gauss can produce sufficient anisotropies.

It seems likely, therefore, that only hydrodynamic effects can produce the required anisotropy. Convective instabilities of the type discussed by Smarr *et al.*⁶ tend to transport matter from the interior of the core into outer layers. Such instabilities could therefore transport trapped neutrino gases outwards toward the neutrinosphere where relatively short diffusion paths would permit rapid escape from the stellar core. Two basic requirements of such a flow, however, would be effective transport to the neutrinosphere and flow-pattern stability over periods of ~ 10 s. Presently available calculations by Smarr *et al.*⁶ tend to indicate smaller-scale, shorter-lived flow patterns than we would require. Another possibility, raised by Spiegel and Prendergast¹¹ in the context of photon transport in upper main sequence stars close to the Eddington luminosity limit, is that high neutrino fluxes can produce a porous fluidized bed just beneath the neutrinosphere.

We note that the neutrino luminosities computed on the basis of the observed neutrino flux are high—perhaps only an order of magnitude or even less below the neutrino Eddington luminosity. Neutrino escape may then be facilitated by the formation of fingers of neutrino fluid or gases that puncture the neutrinosphere. Once such an escape path has been opened, the outward rush of neutrinos through such a duct to the surface may prevent its collapse. The neutrino flow would fill such ducts in the same way that Spiegel and Prendergast's photon gas was envisaged as filling their convective bubbles. Currently, it is not known whether such ducts can persist for as long as 10 s.

An observer viewing a vent or nozzle through the surface of a rapidly rotating core would, therefore, detect particularly strong neutrino emission directed toward him. The degree of beaming would depend on the detailed structure of the neutrino flow.

If the apparent periodicity in the arrival times of neutrinos measured by the Kamiokande detector is real, we expect that a pulsar should be found with a period of 8.9 ms. This may be a useful guide to observers searching for the first faint indications of a pulse, at early epochs, as the supernova ejecta currently hiding the interior are beginning to thin out.

We also note that the timing may be used to establish a stronger upper limit on travel-time effects due to a finite neutrino mass. As energies are observed over several tens of MeV, an upper limit to the neutrino mass, averaged over all flavours detected, is only ≈ 0.2 eV/ c^2 for 1-ms residuals. This would improve on recently estimated upper bounds^{12,13}.

We are grateful to Dr Y. Totsuka from the Kamiokande collaboration for discussing the details of the team's detector, for sending us precise neutrino arrival times and for elucidating the results obtained. We also wish to thank the IMB collaboration for providing us with their unpublished precise timing results. One of us (M.H.) wishes to thank the Alexander von Humboldt foundation for support and the Max Planck Institute for Radioastronomy in Bonn for hospitality. P.B. wishes to thank Dr G. Börner and Dipl. Phys. W. Krülls for discussions on the subject of this paper. I.M.W. wishes to acknowledge support from the NSF, NASA and the Alfred P. Sloan Foundation.

Received 29 April, accepted 29 June 1987.

1. Hirata, K. *et al.* *Phys. Rev. Lett.* **58**, 1490 (1987).
2. Bionta, R. M. *et al.*, *Phys. Rev. Lett.* **58**, 1494 (1987).
3. Woosley, S. E. & Weaver, T. A. *A. Rev. Astr. Astrophys.* **24**, 205 (1986).
4. LeBlanc, J. M. & Wilson, J. R. *Astrophys. J.* **161**, 541 (1970).
5. Müller, E. & Hillebrandt, W. *Astr. Astrophys.* **80**, 147 (1979).
6. Smarr, L. *et al.* *Astrophys. J.* **246**, 515 (1981).

7. Meier, D. L. *et al.* *Astrophys. J.* **204**, 869 (1976).

8. Epstein, R. *Astrophys. J.* **219**, L39 (1978).

9. Shapiro, S. & Teukolsky, S. *Black Holes, White Dwarfs and Neutron Stars* 529 (Wiley, New York, 1983).

10. Tassoul, J.-L. *Theory of Rotating Stars* (Princeton University Press, 1978).

11. Spiegel, E. A. & Prendergast, K. H. *Comm. Astrophys. Space Sci.* **5**, 43 (1973).

12. Bethe, H. *Phys. Rev. Lett.* **56**, 1305 (1986).

13. Bahcall, J. N. & Glashow, S. L. *Nature* **326**, 476 (1987).

Orbital eccentricity in classical novae

D. A. Edwards & J. E. Pringle*

Lick Observatory, Board of Studies in Astronomy and Astrophysics,
University of California, Santa Cruz, California 95064, USA

* Permanent address: Institute of Astronomy, Madingley Road,
Cambridge CB3 0HA, UK

The classical nova outburst is caused by a thermonuclear explosion on the surface of a white dwarf^{1,2}. The explosion ignites at the base of an envelope of hydrogen-rich material which has been accreted from a binary companion star. It has long been evident, however, that in the explosion most of the envelope is not consumed by nuclear burning, but is ejected from the system. This is because to consume the envelope mass ($\sim 10^{-4} M_{\odot}$) at the maximum possible rate (with luminosity equal to the Eddington limit) would take ~ 300 years, whereas classical novae return to their pre-outburst brightness within ~ 10 years (ref. 3). Rough confirmation of this is given by the fact that the masses of shells of material observed to be ejected from novae are in the range 10^{-3} to $10^{-5} M_{\odot}$ (ref. 4). Here we consider the effect on the orbital parameters of a classical nova of the ejection of mass during the nova explosion. The most easily observable consequence is the generation of a small eccentricity in the orbit which leads to a luminosity modulation at a period just longer than the orbital period. Observation of such an effect, would have implications not just for interpreting the dynamics of the explosion but also for measuring the secular effect of tidal interaction after the outburst.

There has been considerable speculation as to how this ejection takes place. The initial force of the explosion is capable of ejecting at least some material in a dynamical fashion, and numerical hydrodynamical calculations have demonstrated that up to about a half of the envelope mass can be ejected in this way⁵. The ejection of the remainder of the envelope appears to take place on a longer timescale in the form of a substantial wind, driven either by radiation pressure⁶ or by interaction of the extended atmosphere of this white dwarf with its binary companion^{3,7}. It seems likely that the fraction of the envelope which is ejected promptly depends on the speed class of the nova.

The rapid ejection of matter from the white dwarf has dynamical consequences for the orbit of the binary, which in turn leads to observational effects. For the sake of argument we consider a nova system to consist of a white dwarf of mass $M_1 = 1 M_{\odot}$, and a secondary star of mass $M_2 = 0.6 M_{\odot}$, effective surface temperature $T_2 = 4,500$ K and equivalent volume radius $R_2 = 3.7 \times 10^{10}$ cm. We assume the secondary fills its Roche lobe, so that the binary separation is $a = 1.1 \times 10^{11}$ cm and the period is $P = 4.4$ h. Before and well after the nova explosion, the main optical light from the system is due to the accretion luminosity caused by a steady flow of mass from the secondary onto the white dwarf. If before outburst the mass transfer rate is $\dot{M} = 10^{-8} M_{\odot} \text{ yr}^{-1}$, then at that time the secondary overfills its Roche lobe by an amount $\Delta R = 7.2 \times 10^7$ cm (refs 8,9).

An ejection of mass from the binary affects the orbit in such a way as to produce two distinct observable consequences. First,

if the mass is ejected promptly from the white dwarf, the binary separation increases^{10,11} by $\Delta a \sim a \Delta M / (M_1 + M_2)$. For $\Delta M = 10^{-4} M_{\odot}$ and the above parameters this leads to a decrease in $\dot{M}$ by about 9%. This is consistent with the observation¹² that the pre- and post-eruption luminosities (due to accretion) are the same to within observational error. Moreover, if the ejection is not prompt but takes place slowly enough that the matter can interact tidally with the secondary star as it leaves the system^{3,7}, then Δa is less than the estimate given above, and may even be negative, leading to an increase in $\dot{M}$. There are also other effects, such as direct heating of the secondary by the hot white dwarf remnant¹¹ which may affect the post outburst mass transfer rate. Thus a change in mass transfer rate post outburst even if observed is not readily interpretable.

Second, however, a prompt ejection of mass from the primary leads to an eccentricity e in the binary orbit. If the timescale on which the mass ΔM is ejected is t_{ej} , then the eccentricity induced is¹³

$$e \approx \frac{\Delta M}{M_1} \times \min \{1, P/2\pi t_{ej}\}$$

For the parameters given above, the maximum eccentricity generated by $\Delta M = 10^{-4} M_{\odot}$ is $e \approx 10^{-4}$. Even such a small orbital eccentricity has observable consequences^{9,14} as it leads to a regular variation in mass transfer rate, and hence of system brightness. For $e \approx 10^{-4}$ the overall variation in $\dot{M}$ is 31%, and is 1% even for $e \approx 3 \times 10^{-6}$. We note that since the variation takes place on a period which is slightly longer than (by a few per cent) but not equal to the orbital period¹⁴, the effect of such an eccentricity should be distinguishable from the orbital variation. Most of the eccentricity is generated by the amount of mass ejected dynamically in the explosion, for which $t_{ej} \ll P$, and little is generated by the ensuing slow mass loss. Thus observation of the effect of such an eccentricity gives a direct measurement of the mass ejected promptly. We should also note that there are processes acting which serve to decrease the eccentricity. In particular the interaction of the wind with the secondary star mentioned above^{3,7}, and simple tidal forces on the secondary¹⁰ both act in the direction of decreasing the eccentricity. The timescales of both processes are quite uncertain. The wind interaction (if it occurs) is likely to act on a timescale comparable to the ejection of the envelope as a whole. The usual tidal effect is not likely to be rapid. For such a small eccentricity even for a lobe filling star, the tides do not induce supersonic motions, and the slow convective turnover times in the secondary lead to long tidal timescales¹⁵. Thus, the observation of an eccentricity in the orbit is not only of importance in understanding the dynamics of the nova explosion, but observation of its secular decrease would give the first direct measurement of the timescale on which tides act in a binary star system.

This work was carried out under the auspices of a special NASA astrophysics theory program which supports a joint centre for star formation studies at NASA-Ames Research Center, UC Berkeley, and UC Santa Cruz. We gratefully acknowledge support from NATO (D.A.E.) and the Fulbright Commission (J.E.P.).

This is Lick Observatory Contribution No. 445.

Received 20 May; accepted 16 June 1987.

- Gallagher, J. S. & Starrfield, S. A. *Rev. Astr. Astrophys.* **16**, 171–214 (1978).
- Truran, J. W. in *Essays in Nuclear Astrophysics* (ed. Barnes, C.) 467–495 (Cambridge University Press, 1985).
- MacDonald, J., Fugimoto, M. Y. & Truran, J. W. *Astrophys. J.* **294**, 263–270 (1985).
- McLaughlin, D. B. in *Stars and Stellar Systems* Vol. VI, (ed. Greenstein, J. L.) 585–653 (University of Chicago Press, 1960).
- Starrfield, S., Truran, J. W. & Sparks, W. M. *Astrophys. J.* **226**, 186–202 (1978).
- Bath, G. T. & Shaviv, G. *Mon. Not. R. astr. Soc.* **175**, 305–322 (1976).

- MacDonald, J. *Mon. Not. R. astr. Soc.* **191**, 933–949 (1980).
- Paczynski, B. & Sienkiewicz, R. *Acta. astr.* **22**, 73–92 (1972).
- Edwards, D. A. & Pringle, J. E. *Mon. Not. R. astr. Soc.* (in the press).
- Pringle, J. E. & Wade, R. A. (eds) *Interacting Binary Stars* (Cambridge University Press, 1985).
- Shara, M. M., Livio, M., Moffatt, A. F. J. & Orio, M. *Astrophys. J.* (in the press).
- Robinson, E. L. *Astr. J.* **80**, 515–524 (1975).
- Savedoff, M. P. & Vila, S. *Astr. J.* **69**, 242–245 (1964).
- Papaloizou, J. C. B. & Pringle, J. E. *Mon. Not. R. astr. Soc.* **189**, 293–297 (1979).
- Campbell, C. G. & Papaloizou, J. C. B. *Mon. Not. R. astr. Soc.* **204**, 433–447 (1983).

Comet Bowell at record heliocentric distance

Karen J. Meech & David Jewitt

Department of Earth, Atmospheric and Planetary Sciences,
Massachusetts Institute of Technology, Cambridge,
Massachusetts 02139, USA

Comets probably contain the most primordial material in the Solar System, and those which are passing through the inner Solar System for the first time are likely to be the most pristine of all. Comet Bowell (1982 I = 1980b) was discovered in 1980 at the unusually large heliocentric distance of $R = 7.3$ AU (ref. 1). The small reciprocal of the original orbital semi-major axis, $(1/a) = 30 \times 10^{-6} \text{ AU}^{-1}$, suggests that this comet is dynamically new in the sense that it has made few, if any, previous passages through the inner Solar System². Because prolonged solar irradiation may alter the properties of short-period comets, it is of great interest to see if 1982 I exhibits properties which are significantly different from those of the periodic comets. We present new observations of comet Bowell at the record distance $R = 13.6$ AU. An extended coma is present, the size of which is consistent with the same slow expansion rate $v \approx 1 \text{ m s}^{-1}$ detected around perihelion (March 1982 and $R \approx 3.36$ AU). The cross-section of the solid grains within the central 10 arc s of the coma has decreased by over an order of magnitude since 1980–1984, which indicates that the coma production is declining. The decline began near $R \approx 10$ AU, the same distance at which production began on the pre-perihelion leg. The coma at $R \leq 10$ AU may be formed by sublimation of CO_2 or an ice of similar volatility from the nucleus.

We briefly summarize the major characteristics of 1982 I previously determined from observations around perihelion (1980–84). The comet is known to exhibit substantial activity at distances beyond which H_2O sublimation is expected to be significant ($R \geq 6$ AU). Its lightcurve is quite different from that of comet Halley, in which the activity is controlled largely by H_2O sublimation³. There has been one reported detection of H_2O ice in the near infrared spectrum of this comet⁴, however the identification is controversial⁵. Other near infrared spectra of the comet show no evidence for H_2O ice^{5,6}, and water-ice sublimation models are unable to reproduce the lightcurve⁷. More volatile ices, typified by N_2 , provide a better (but not perfect) match to the near-perihelion photometry⁷. The best match to the photometry is provided by an inert coma, but the solid-grain coma was found to expand at a very low velocity $v = 0.9 \pm 0.2 \text{ m s}^{-1}$ (refs 7, 8). Estimates of the ratio of the radiation pressure acceleration to the gravitational acceleration of particles in the tail indicate that comet Bowell has an unusual particle size distribution, with a deficiency of particles of radii $a \leq 300\text{--}400 \mu\text{m}$ (ref. 8). Given the unusual properties of the comet and the likelihood that H_2O sublimation was not responsible for its activity at $R \geq 6$ AU, we were encouraged to attempt to observe this comet at extremely large R to obtain information

about the physical processes controlling the activity on the nucleus.

The present observations were obtained with the Kitt Peak 2.1-m telescope and an 800×800 Texas Instruments charge-coupled device (CCD), using an ephemeris provided by B. G. Marsden (personal communication). Four images of 900 s each were taken through a Mould R filter on 3 November 1986 UT ($R = 13.6$ AU), with the telescope tracked at sidereal rate (see Table 1 for observational parameters). The comet appeared as a diffuse patch of extremely low surface brightness which clearly showed the expected motion between exposures. A summed image of comet Bowell, made by shifting the individual images according to the ephemeris motion (giving an effective exposure of 1 hour), is shown in Fig. 1a. A line diagram of the same image showing the location of the comet with respect to trailed stars is given in Fig. 1b. Unmarked diffuse features in Fig. 1a are faint field galaxies. We also detected comet Bowell (at $R = 11.0$ AU) on 21 September 1986 UT (see Table 1) using the Kitt Peak 4-m telescope with the same CCD at the prime focus.

The brightness profile of the comet at $R = 13.6$ AU is shown in Fig. 2. The brightness was measured as a function of projected distance from the nucleus along a line drawn perpendicular to the projected orbit. From Fig. 2 we obtain an estimate of the lower limit of the radius of the coma as $P_{13.6 \text{ AU}} = 15 \pm 2 \text{ arc s}$. A similar measurement from the 4-m image yields $P_{11.0 \text{ AU}} = 36 \pm 3 \text{ arc s}$. These and earlier radius measurements^{7,9} are plotted against the Julian Day (JD) number in Fig. 3. The large error bars assigned to measurements from ref. 9 reflect the fact that the radii were obtained from profiles which were computed in circular annuli centred on the nucleus instead of from cuts perpendicular to the orbital plane (inclination to the ecliptic $\approx 1.7^\circ$) and hence may contain the effects of radiation pressure. It is clear from Fig. 3 that the coma size is not symmetric with respect to perihelion (represented by a vertical dashed line) as would be expected for a normal coma¹⁰. Instead, a least squares fit to the data in Fig. 3 (excluding the lower limit at JD = 2,446,737.6) gives the constant coma expansion velocity $v = 1.1 \pm 0.2 \text{ m s}^{-1}$. The fit is shown as the solid line in Fig. 3. There is good agreement with the previous best fit⁷, $v = 0.9 \pm 0.2 \text{ m s}^{-1}$, shown as a dashed line in Fig. 3. The inferred time of coma formation (the time at which the coma radius was zero) is $\text{JD}_0 = 2,443,934 \pm 900$ days ($R \approx 10$ AU), also in agreement with previous determinations^{7,8}.

The surface brightness of the coma of comet Bowell is $<1\%$ of the surface brightness of the night sky in the R filter (see Fig. 2), making accurate photometry extremely difficult. Our best estimate of the magnitude within an aperture of radius 5 arc s is $m_R(5'') = 21.7 \pm 1.0 \text{ mag}$ at $R = 13.6$ AU, where the large uncertainty is due to sky subtraction error and to variable extinction. At $R = 11.0$ AU the magnitude within an aperture of radius 5 arc s is $m_R(5.0'') = 20.0 \pm 0.2$. As a matter of interest, a crude estimate of the total brightness of the coma can be computed. We find $m_R \approx 17.9$ at $R = 11.0$ AU and $m_R \approx 20.5$ at $R = 13.6$ AU; even at 13.6 AU the comet is not particularly faint, only its diffuseness makes it hard to detect. In previous observations, the grain cross-section in a diaphragm of 5 arc s radius remained

Table 1 Observational parameters

JD (−2,440,000)	UT† (mid)	Exp‡ (s)	Tel (m)	Filter	χ airmass	α (1950)	δ (1950)	r (AU)¶	Δ (AU)#	Phase (deg)
6,737.6373	3:17:41	900	2.1	R	1.32	00:17:57.0*	00:06:07	13.562	12.745	2.41
6,737.6734	4:09:42	900	2.1	R	1.23	00:17:56.0	00:06:07	13.562	12.745	2.41
6,737.7373	5:41:41	900	2.1	R	1.20	00:17:56.4	00:05:58	13.562	12.745	2.41
6,737.7726	6:32:35	900	2.1	R	1.26	00:17:55.5	00:05:55	13.462	12.745	2.41
6,329.9444	10:40:00	900	4	R	1.84	00:01:32.6†	−01:38:46	11.006	10.003	0.20

* Positions as measured at the telescope by offsetting from nearby SAO stars. The positions are accurate to approximately ± 5 arc s.

† Positions accurate to approximately ± 30 arc s.

‡ UT midtime of integration; § exposure duration in seconds; || telescope diameter in m; ¶ heliocentric distance in AU; # geocentric distance.

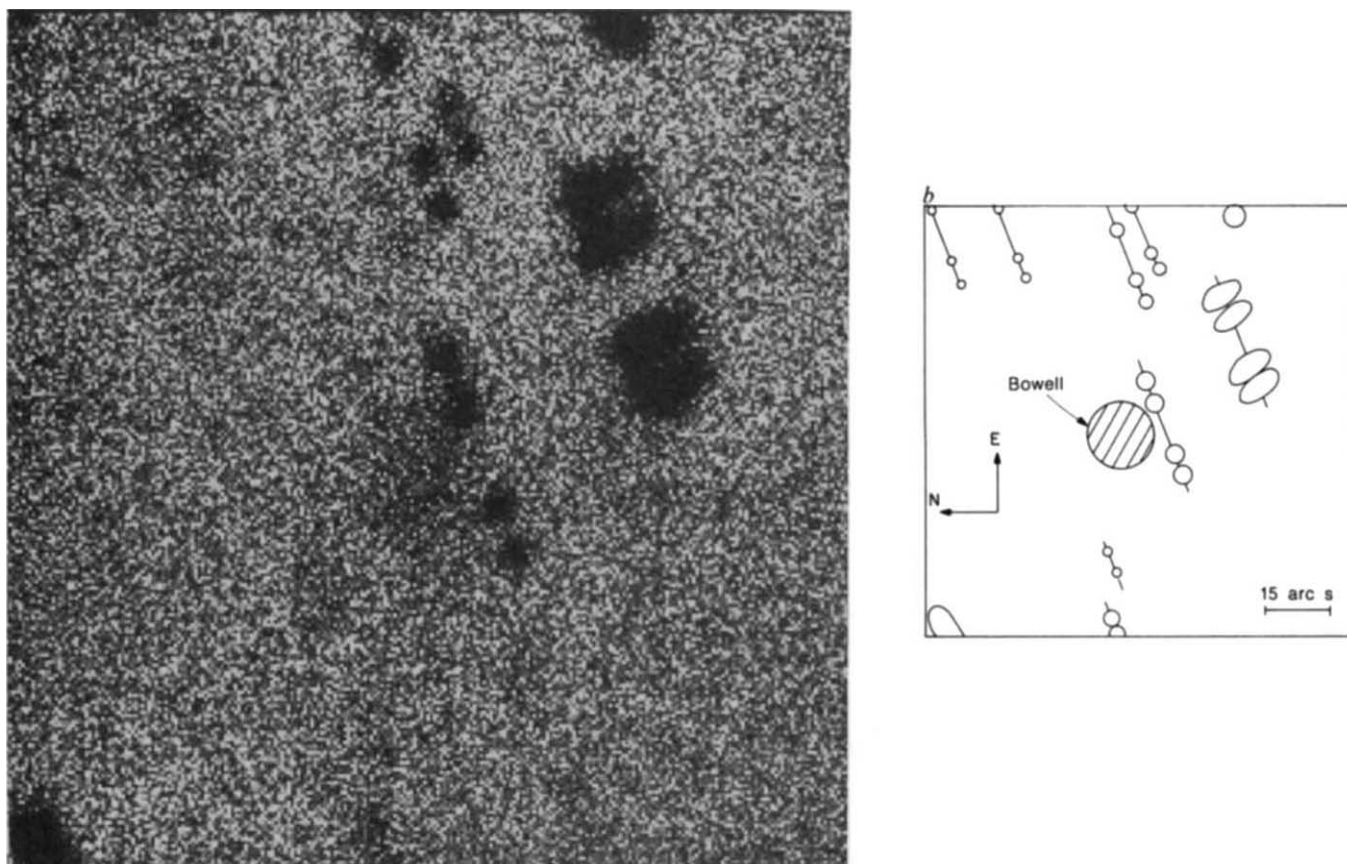


Fig. 1 *a*, CCD image of comet Bowell at $R = 13.6$ AU, taken 3 November 1986 UT. Four 900-s images were shifted according to the motion of the comet and a median filtered cubic spline function was fitted to the sky background near the comet and subtracted to produce this 1 h effective exposure. North is to the left and East to the top. The comet appears diffuse, with a total integrated magnitude within a diaphragm of 15 arc s radius of $m_R = 20.5 \pm 1.0$. *b*, Line diagram of Fig. 1*a* showing the location of the brighter field stars and galaxies with respect to the comet. The circle indicates the position but not the size of comet Bowell. A scale bar is shown for reference.

constant⁷ at $C = (1.6 \pm 0.3) \times 10^9 \text{ m}^2$ (where $C = 2.24 \times 10^{22} R^2 \Delta \pi 10^{0.4(m_R(\text{sun}) - m_R(5''))}/p$; and $p \approx 0.06$ (ref. 11) is the geometric albedo of the grains). The new photometry shows that the total cross-section has decreased by about an order of magnitude to $C = (1.8 \pm 0.3) \times 10^8 \text{ m}^2$ at $R = 11.0$ AU and $C = 7 \times 10^7 \text{ m}^2$ (within a factor of 2) at $R = 13.6$ AU. The quoted errors reflect not only the statistical uncertainty, but also the uncertainty in assuming a brightness profile which falls off as the inverse of the projected distance from the nucleus. Thus, the photometry suggests that the coma production is declining, and that the decline began at about the same critical heliocentric distance at which the activity began ($R \approx 10$ AU).

Numerous hypotheses have been put forward to explain the existence of cometary coma at large R , including (1) sublimation from small amounts of 'super-volatiles' such as CO_2 or C_2H_4 at large R (ref. 12); (2) gravitational perturbations by the sun on a dormant coma which is a remnant of the comet's formation⁸, (3) electrostatic snap-off⁷; or (4) chemical instabilities¹³ and phase transitions in the ice at low temperatures^{14,15}. Of all these processes, the one which provides the most natural explanation for the appearance of a coma near $R = 10$ –11 AU is the sublimation of material more volatile than H_2O . In particular, we are intrigued that near $R = 10$ AU, the momentum carried by sublimating CO_2 would just be sufficient to drag submillimetre particles into the coma from a nucleus of a few km radius.

The existence of CO_2 has been inferred in other comets. Comet Schwassmann–Wachmann 1 orbits the sun between 5–7 AU and has dramatic outbursts in which CO^+ has been detected¹⁶. It has been suggested that the sublimation of CO_2 or CO which

is suddenly exposed to the surface is directly related to the comet's activity¹⁷. In addition, CO_2^+ is frequently observed in cometary plasma tails. Recent observations of sudden increases in the column density of CO_2^+ ions in the tail of comet Halley suggest that CO_2 probably plays a role in the outbursts¹⁸. The birth and probable death of the coma of 1982 I near $R = 10$ AU, the

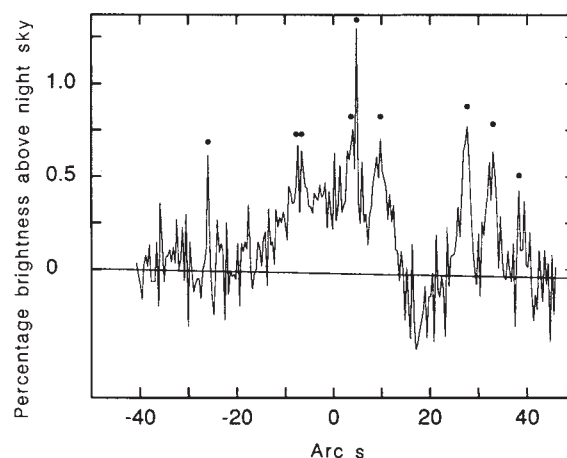


Fig. 2 Surface brightness profile E-W through comet Bowell at $R = 13.6$ AU. The profile is the mean surface brightness within a band of width 7.6 arc s centred on the comet. The level of the sky background is shown by a straight line. Several stars contaminate the profile; these are indicated by dots.

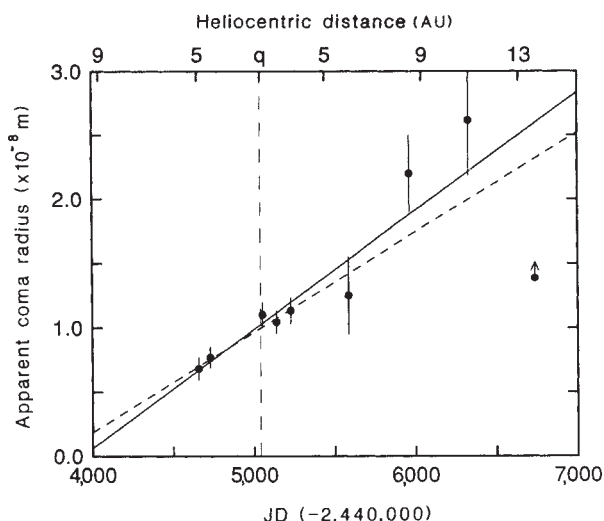


Fig. 3 Plot of the coma radius versus JD and heliocentric distance. Perihelion (at $q = 3.364$ AU) is marked with a vertical line. The least squares fit to the radius before 1984 is shown as the dashed line, the fit including the new data is shown as the solid line. The velocity of expansion is $v = 1.1 \pm 0.2 \text{ m s}^{-1}$. The data point at JD = 2,446,737.6 represents a lower limit and is not included in the fit.

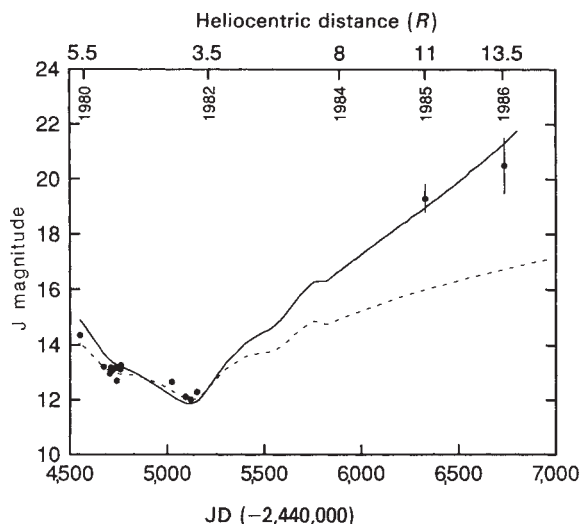


Fig. 4 J magnitudes with a 5 arc s radius diaphragm are plotted versus JD. The heliocentric distances and dates are marked at the top of the graph. The CO_2 sublimation model described in the text (solid line) is compared to the 'constant cross-section' model (dashed line).

existence of submillimetre grains in the coma and the ability of CO_2 sublimation to drag large grains from the coma at this distance all encourage a closer look at the CO_2 sublimation hypothesis.

To evaluate the feasibility of CO_2 sublimation for comet Bowell, we compare all of the available broadband photometry corrected to a 5 arc s radius diaphragm (the present photometry and that summarized in ref. 7) to a simple CO_2 sublimation model. The CCD R magnitudes (this paper) have been converted to the J bandpass (used in ref. 7) assuming $(R - J)_{\text{sun}} = 0.57$ (ref. 19). The model assumes equilibrium sublimation from a nucleus of negligible conductivity and is analogous to the model successfully applied to comet Halley (see ref. 3 for a complete description). Unfortunately, in the case of comet Bowell, relatively few of the input parameters are well known. What is known is an upper limit to the product of the $0.65 \mu\text{m}$ geometric albedo and cross-section of the nucleus⁷, $C_n \leq 6 \times 10^6 \text{ m}^2$; in addition, the analysis of the tail morphology provides a good estimate of the mean particle size⁸. Observations^{7,11} also indicate that the albedo of the grains is low (we adopt $p \approx 0.06$). Figure 4 presents a comparison of the CO_2 model with the J magnitude in a 5 arc s radius diaphragm. It is interesting that the sublimating CO_2 model is able to fit the broadband photometry quite well. Also shown in Fig. 4 is a 'constant cross-section' model which shows the brightness variation that an extended source would have within a fixed diaphragm as the geometric factors changed (dashed line). This model fits the photometry well near perihelion, but is inadequate at larger distances.

A strict analogy with Halley would suggest that CO and not CO_2 should be the dominant volatile since the abundance of CO relative to H_2O is $\sim 20\%$ in Halley²⁰, whereas the abundance of CO_2 relative to H_2O is only $\sim 3.5\%$ ²¹. Models for the much more volatile CO are compatible with the comet Bowell data only for very high-grain/nucleus albedos; however, high grain albedos are inconsistent with measured values^{7,11}.

The CO_2 sublimation model predicts that the terminal velocity of $300\text{-}\mu\text{m}$ grains in the outflowing gas should be of the order of 20 m s^{-1} at perihelion ($R \approx 3.36$ AU), decreasing as R increases. But the observed coma expansion velocity is $v \approx 1 \text{ m s}^{-1}$, near the expected escape speed from the nucleus. It is difficult to reconcile the slow coma expansion with the expected 20 m s^{-1} terminal speed of the grains. One possible solution to this dilemma could be offered if the mass of dust ejected into

the coma at high speed near perihelion were negligible compared to the total mass of dust present in the coma. The mass of dust in the observing diaphragm is equal to the integral of the mass loss rate over the time it takes the grains to cross the diaphragm. Therefore, if the comet experiences high ejection velocities and high mass loss for a relatively short time, it is possible that the total amount of material ejected at high velocity would represent only a small fraction of the total mass of dust in the coma and would hence be undetected. Unfortunately, our calculations indicate that the high terminal velocity is sustained for a period of time long enough that the high-velocity dust grains should have been readily detectable. Although the CO_2 model has difficulty explaining the observed low velocity of the grains, near perihelion the initiation of the coma near $R = 10\text{--}11$ AU and the apparent turn off at the same distance most strongly suggest sublimation of a material with the volatility of CO_2 .

For grains moving at 1 m s^{-1} , the diaphragm crossing time is of the order of 1 yr. The comet has now been observed for more than 6 yr, significantly longer than the diaphragm-crossing time. The long diaphragm-crossing time effectively provides a long time constant for photometric changes. The nearly constant coma cross-section at $R \leq 10$ AU suggests that the mechanism producing the coma has been continuous and was not a single impulsive ejection at large R . The suggestion of an ice phase transition¹⁵ thus seems unlikely on the basis that the coma was active both before and after perihelion whereas a phase transition would occur on the inbound journey as the comet first began to heat and would not naturally produce continued activity at the same distance post-perihelion.

The most distant recorded comet before Bowell, comet Stearns (1927 IV) at $R = 11.5$ AU (ref. 22), was also active at very large distances from the sun. Analysing photographic observations²²⁻²⁴ from 1927-32, we find that comet Stearns was in some respects similar to comet Bowell (K.J.M., in preparation). Each comet had an extensive coma to $R \approx 10\text{--}11$ AU and both showed peculiar narrow tails. It will be crucial to obtain observations of comets which are active at very large heliocentric distances, of the type represented by comets Bowell and Stearns, to understand the source of the activity and to compare them with the short-period comets.

We are both Visiting Astronomers at the Kitt Peak National Observatory. We thank the efficient Kitt Peak staff for their excellent support, in particular we thank our telescope operator,

George Will. This work was supported in part by the NSF and NASA.

Received 28 January; accepted 30 April 1987.

1. Bowell, E. *IAU Circ.* No. 3465 (1980).
2. Everhart, E. & Marsden, B. G. *Astr. J.* **93**, 753-754 (1987).
3. Meech, K. J., Jewitt, D. & Ricker, G. R. *Icarus* **66**, 561-574 (1986).
4. Campins, H., Rieke, G. H. & Lebofsky, M. J. *Nature* **301**, 405-406 (1983).
5. A'Hearn, M. F., Dwek, E. & Tokunaga, A. T. *Astrophys. J.* **282**, 803-806 (1984).
6. Jewitt, D. C., Soifer, B. T., Neugebauer, G., Matthews, K. & Danielson, G. E. *Astr. J.* **87**, 1854-1866 (1982).
7. Jewitt, D. *Icarus* **60**, 373-385 (1984).
8. Sekanina, Z. *Astr. J.* **87**, 161-169 (1982).
9. Baum, W. A., Kreidl, T. J. & Schleicher, D. G. *Bull. Am. astr. Soc.* **18**, 794 (1986).
10. Mendis, D. A. & Ip, W.-H. *Astrophys. Space Sci.* **39**, 335-385 (1976).
11. Hanner, M. S., Veeder, G. J. & Matson, D. L. *Bull. Am. astr. Soc.* **3**, 705-706 (1981).
12. Mukai, T. *Astr. Astrophys.* **164**, 397-407 (1986).
13. Donn, B. & Urey, H. C. *Astrophys. J.* **123**, 339-342 (1956).
14. Smoluchowski, R. *Astrophys. J.* **244**, L31-L34 (1981).
15. Hanner, M. S. & Campins, H. *Icarus* **67**, 51-62 (1986).
16. Cochran, A. L., Barker, E. S. & Cochran, W. D. *Astr. J.* **85**, 474-477 (1980).
17. Cowan, J. J. & A'Hearn, M. F. *Icarus* **50**, 53-62 (1982).
18. Feldman, P. D. *et al.* *Nature* **324**, 433-436 (1986).
19. Johnson, H. L. in *A. Rev. Astr. Astrophys.* Vol. 4 (eds Goldberg, L., Layzer, D. & Phillips, J. G.) 193-205 (Annual Reviews Inc., California, 1966).
20. Woods, T. N., Feldman, P. D., Dymond, K. F. & Sahnou, D. J. *Nature* **324**, 436-438 (1986).
21. Krankowsky, D. *et al.* *Nature* **321**, 326-329 (1986).
22. Van Biesbroeck, G. *Astr. J.* **42**, 25-32 (1972).
23. Burton, H. E. *Astr. J.* **38**, 35-36 (1927).
24. Van Biesbroeck, G. *Astr. J.* **40**, 51-60 (1930).

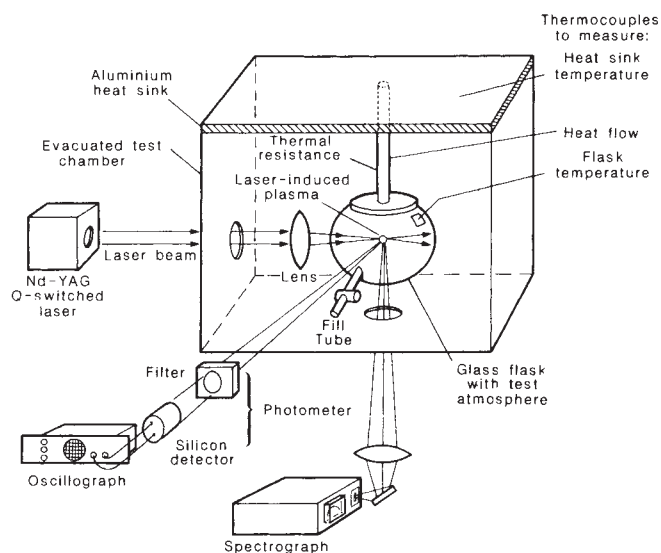


Fig. 1 Apparatus to measure optical efficiencies and spectral irradiance.

Optical efficiencies of lightning in planetary atmospheres

W. J. Borucki & C. P. McKay

NASA Ames Research Center, Moffett Field, California 94035, USA

Spacecraft observations show that the presence of lightning activity is not confined to the terrestrial atmosphere, but is also found in the atmospheres of Venus¹, Jupiter² and Saturn³. Lightning activity may also occur in Titan's thick atmosphere. Calculations⁴⁻⁶ show that lightning produces a significant fraction of the nitric oxide that reacts with the ozone and chlorine compounds in the terrestrial stratosphere. In the atmosphere of the primordial Earth, lightning could have been the major source of many of the molecules required for the formation of life⁷. To determine the effects of lightning activity in the atmospheres of other planets from spacecraft images requires a knowledge of the optical properties of the lightning discharge. Here we report the first simulations of lightning in planetary atmospheres by laser-induced plasmas. These simulations show that the fraction of the energy in lightning discharge channels that is radiated in the visible spectrum is similar for Earth, Venus and Titan, but quite different for Jupiter. One implication of our results is that the amount of trace gases produced by lightning in the jovian atmosphere must be larger than previously estimated.

It is expected that the trace-gas production rate by lightning will be proportional to the rate of energy dissipated by lightning activity. The energy dissipation rate can be estimated from the average optical energy dissipated per flash and the global flash rate if the optical efficiency is known. The optical energy dissipated per flash and the global flash rate are obtained from calibrated images taken by spacecraft. The optical efficiency is defined to be the ratio of the optical energy to the total energy deposited in the discharge column. Thus the determination of the optical efficiency is critical to the estimation of the trace-gas production rate by lightning.

Because of experimental difficulties, only estimates of the optical efficiency of terrestrial lightning have been reported previously. Connor⁸ used a spectrograph with calibrated film and electric field data to obtain an estimate of 7×10^{-3} for the optical efficiency of terrestrial lightning. The uncertainty⁴ in his estimate is at least a factor of four. To estimate more accurately

the total energy dissipated in the discharge and the optical energy radiated by the discharge column, Krider *et al.*⁹ used a silicon diode photometer to observe a 4-m spark in air. The electrical energy dissipated was determined in two ways: (1) integrating the product of the current and voltage with time; and (2) subtracting the energy stored in the impulse generator before and after the discharge. The agreement between the two measurements of the energy dissipation was within 40%. They conducted time-resolved spectroscopic studies of both the spark and lightning and found that both sources had similar spectra and should have similar optical efficiencies. They obtained an average optical efficiency of 3.8×10^{-3} .

Recent measurements¹⁰ in our laboratory have shown that the emission spectrum of terrestrial lightning can be simulated by laser-induced plasmas (LIP). Because such plasmas do not contact any electrodes, no contaminant spectra are present to interfere with the spectrum of the test gas and thus it is not necessary to use metre-long discharges to obtain adequate spectral purity of the optical radiation. A few cubic centimetres of gas are sufficient to produce the LIP and measure the optical efficiency. Hence, it is feasible to work with flammable gases like hydrogen and methane that are important components of planetary atmospheres.

Our measurements were made with a Q-switched Nd-YAG laser, a silicon diode detector, and a specially developed powermeter that measured the absorbed laser radiation.

Figure 1 is a sketch of the experimental apparatus. The flask with the test atmosphere is shown in the central portion of this figure. A 5-ns pulse of parallel light from the laser enters from the left, passes through a condensing lens to a focus at the centre of the flask. At the lens focus, the laser radiation causes gas breakdown and the formation of a high-temperature plasma. Any portion of the laser beam that is transmitted past the LIP exits through a window at the right-hand side of the test chamber. Laser radiation scattered from the plasma also passes out of the flask. Consequently, only power absorbed from the laser beam by the LIP is measured by the powermeter. The photometer measures the optical power as a function of time and records these data on an oscilloscope. To exclude scattered laser radiation from the photometer, an optical filter that blocks the $1.1\text{-}\mu\text{m}$ laser radiation is used in front of the detector. The spectrograph verifies the purity of the test gases used.

To determine the total energy deposited by the laser in the LIP, the flask is placed in an evacuated chamber and the steady-state heat production in the flask is determined by measuring the temperature gradient at equilibrium in the aluminium rod

Table 1 Optical efficiency values

Planetary atmosphere	Excitation method	Optical efficiency	Reference
Earth	Lightning	0.0070	Ref. 8
Earth	4-m spark	0.0038	Ref. 9
Earth	Laser-induced plasma	0.0035	Present work
Venus	Laser-induced plasma	0.0034	Present work
Jupiter	Laser-induced plasma	0.0010	Present work
Titan	Laser-induced plasma	0.0042	Present work

at the top of the flask. The inferred value of the heat flow through the rod is corrected by adding experimentally determined corrections for the convective and radiative heat losses from the flask and the absorption of the laser radiation by the flask walls.

The optical efficiency η , of the LIP is defined to be the ratio of the optical energy E_0 , radiated in the 420–900 nm spectral range divided by the total energy E_t , in the LIP. E_t is determined from the powermeter and the known pulse repetition rate. E_0 is determined from the photometer measurements. From an error analysis of the random and systematic errors, we estimate the total uncertainty in our measurements to be $\pm 32\%$. A more detailed discussion of the experimental procedure and analysis is given in ref. 11.

The gas mixtures used to simulate the atmospheres of Venus, Jupiter and Titan were 96% CO₂:4% N₂, 87% H₂:13% He and 97% N₂:3% CH₄, by volume, respectively. Table 1 lists the values of η obtained for these gas mixtures. The value of η derived from the LIP measurements for the terrestrial atmosphere is close to the value derived for a 4-m spark by Krider *et al.*⁹, but is only about half that derived by Connor⁸ for terrestrial lightning. The cause of this difference is unclear, but might be due to the large uncertainty of his measurement. The excellent match of the two laboratory methods for simulating lightning, and the reasonable agreement of these methods with the field measurement of Connor⁸, support the feasibility of using laboratory simulations to estimate the optical efficiencies of lightning on other Solar System bodies.

The values in Table 1 show that lightning on Venus, Earth and Titan have similar optical efficiencies and that these efficiencies are much higher than the jovian value. The much lower value for jovian lightning is as expected given that the jovian lightning spectrum is nearly devoid of line radiation¹⁰. The significance of the small value of η for jovian lightning is that the previous estimate¹² of the energy dissipation rate of lightning was based on the assumption that the jovian value and terrestrial value of η were the same. Because the measured value is approximately 30% of the previously assumed value, the estimate of the energy dissipation rate must be raised from $4 \times 10^{-4} \text{ W m}^{-2}$ to $1.2 \times 10^{-3} \text{ W m}^{-2}$. At first glance it would appear that the estimate¹³ of the trace-gas production rates of CO, C₂H₂ and HCN by lightning activity should also be raised by a similar factor. But their estimate is based on an analysis of the chemistry induced by the lightning activity that is sensitive to the altitude region in which the lightning occurs. Because a recent analysis¹⁴ of Voyager 1 images indicates that the jovian lightning activity is occurring near the top of the water cloud rather than well above the water cloud as assumed in ref. 13, an improved estimate of the trace-gas production rates must await an effort that includes both effects.

Received 3 February; accepted 28 May 1987.

1. Ksanfomality, L. V., Scarf, F. L. & Taylor, W. W. L. in *Venus* 565–603 (University of Arizona Press, Tucson, 1983).
2. Cook II, A. F., Duxbury, T. C. & Hunt, G. E. *Nature* **280**, 794 (1979).
3. Burns, J. A., Showalter, M. R., Cuzzi, J. N. & Durisen, R. H. *Icarus* **54**, 280–295 (1983).
4. Borucki, W. J. & Chameides, W. L. *Rev. Geophys. Space Phys.* **22**, 363–372 (1984).
5. Bhetanabhotla, M. N., Crowell, B. A., Coucouvinos, A., Hill, R. D. & Rinker, R. G. *Atmos. Envir.* **19**, 1391–1397 (1985).
6. Ko, M. K. W., McElroy, M. B., Weissenstein, D. K. & Sze, N. D. *J. geophys. Res.* **91**, 5395–5404 (1986).

7. Chameides, W. L. & Walker, J. C. G. *Orig. Life* **11**, 291–302 (1981).
8. Connor, T. R. *The 1965 ARPA-AEC Joint Lightning Study at Los Alamos* Vol. 15, LA-3754 (Los Alamos Scientific Lab., Los Alamos, New Mexico, 1967).
9. Krider, E. P., Dawson, G. A. & Uman, M. A. *J. geophys. Res.* **73**, 3335–3339 (1968).
10. Borucki, W. J., McKenzie, R. L., McKay, C. P., Duong, N. D. & Boac, D. S. *Icarus* **64**, 221–232 (1985).
11. Borucki, W. J. *et al. Appl. Opt.* (in the press).
12. Borucki, W. J., Bar-Nun, A., Scarf, F. L., Cook II, A. F. & Hunt, G. E. *Icarus* **52**, 492–502 (1982).
13. Bar-Nun, A. & Podolak, M. *Icarus* **64**, 112–124 (1985).
14. Borucki, W. J. & Williams, M. A. *J. geophys. Res.* **91**, 9893–9903 (1986).

Waves in the atmosphere of Venus

Peter J. Gierasch*

Department of Atmospheric Physics, Clarendon Laboratories, Oxford University, Parks Road, Oxford OX1 3PU, UK

The recent Soviet/French instrumented balloon experiments in the Venus atmosphere have shown that turbulent vertical motions with amplitudes of several metres per second exist within the principal cloud deck at about 50 km elevation¹. The motions are particularly intense over the Aphrodite mountains, and the experimenters suggest that a disturbance forced by surface flow over the mountains reaches to cloud levels^{1,2}. Here I point out that a patch of turbulence stationary over a surface feature can generate a wake-like pattern of horizontally propagating waves which can explain the general visual appearance of Venus, and which can produce a momentum exchange from mid-latitudes to low latitudes, helping to maintain the Venus atmospheric rotation.

The mean zonal flow in the Venus atmosphere increases with height from the surface to about 70-km elevation³, reaching about 100 m s⁻¹. Its direction is the same as that of the solid planet's rotation (retrograde), but about 50 times stronger. The cause of this strong superrotation is not understood⁴. At about 50 km above the surface there is a cloud deck of sulphuric acid droplets⁵ which is bland in appearance in visible light but shows features at ultraviolet wavelengths, as displayed in Fig. 1. Figure 2 shows a sketch of the stability profile in the atmosphere of Venus as measured by the Pioneer Venus descent probe⁶. Although there is some variability between measurements made by different probes the layer between 48- and 55-km elevation always shows small stability. It is probable that the radiative properties of the atmosphere drive this layer toward convection, particularly at low latitudes where solar heating is strongest⁷. The deeper layer between 20 and 30-km elevation also consistently shows small stability. The sandwiched stable layer between these can be a duct for horizontally propagating internal gravity waves. I propose that such waves are produced at low latitudes by flow past a surface-fixed convective disturbance. The energy available to produce waves can be great if convective instability of the basic state can be tapped. Note that the forcing is seated in the convective layer just above the duct, not in it. For effective coupling, the wave modes must have amplitude distributions which extend vertically into the convecting layer. I have numerically tested this point, as reported below.

Vertical penetration is also necessary for wave effects to reach the visible layers just above the convecting region, although the question of visibility raises subtle and unresolved issues. The waves themselves will not be directly visible in the cloud because the vapour pressure over the cloud droplets is extremely small⁴, and they behave as an inert aerosol on short timescales. Consistent with this, observations show the cloud patterns drifting with the velocity of the mean flow⁸. But at certain locations the waves will be subject to nonlinear effects, and will produce irreversible overturning. For example, a favoured location for such overturning might be along caustics, where amplitudes

* Present address: Astronomy Department, Cornell University, Ithaca, New York 14853, USA.

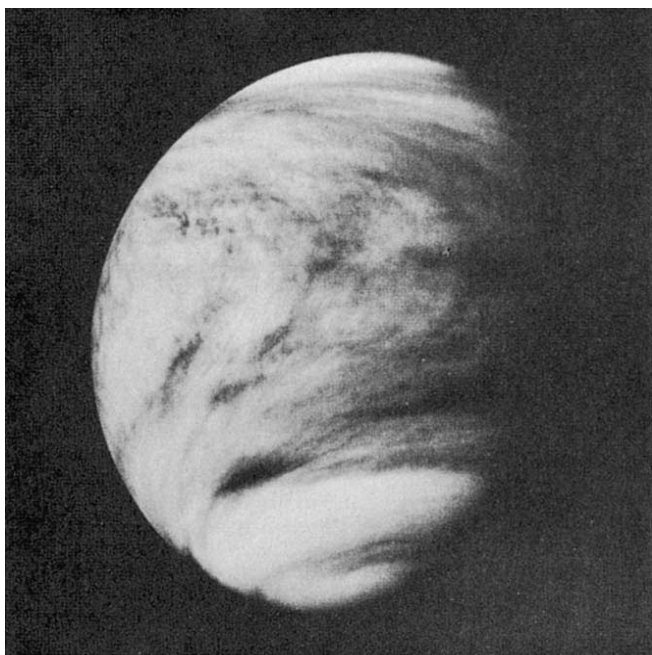


Fig. 1 Pioneer Venus image in ultraviolet light.

become large. Irreversible overturning will bring compositional mixing which can affect the albedo of the cloud, and it is to this we attribute the observed patterns. Thus, the visible features are analogous to foam produced by the breaking of ocean waves, where the long-lived foam drifts with the mean motion of the sea.

The momentum carried by the horizontally propagating waves will be in such a sense as to accelerate the mean flow near the equator and decelerate it at mid- or high latitudes, wherever the waves are absorbed. The direction of the mean flow is westward and the waves therefore propagate eastward relative to the fluid, while at the same time spreading away from the equator. These waves transport momentum in the same direction as their phase propagation, and thus they will deposit eastward momentum at high latitudes. Therefore, in the region where they are generated an impulse in the direction of the mean flow, westward, is imparted. The net effect is thus to transfer angular momentum from high latitudes toward the equator, in the correct sense to accelerate the equatorial regions. These waves therefore may play a key role in maintaining the planet's circulation. They may also provide the first explanation of the magnitude of the superrotation, because as we show below, ducted waves have a maximum propagation velocity only slightly larger than the velocity which now obtains for the mean flow at the level of the duct.

We turn now to a description of the properties of horizontally propagating internal gravity waves in the duct just below the cloud deck. The governing equations are well known and well understood^{9,10}. The stability is measured by the Brunt frequency N , given by

$$N^2 = \frac{g}{T} [T_z - (T_z)_{\text{adiabatic}}] \quad (1)$$

Here g is the acceleration of gravity, $T(z)$ is temperature, and a subscript denotes differentiation. The trapping or propagation of waves in height is determined primarily by the temperature profile (through N^2) and by the zonal wind profile $\bar{u}(z)$. For waves stationary with respect to the surface, the relative velocity between the waves and the wind is about $\bar{u} = 50 \text{ m s}^{-1}$ near the centre of the duct. The Brunt frequency near the middle of the stable layer can be estimated from the data displayed in Fig. 1, and is $\sim 0.01 \text{ s}^{-1}$. The depth of the stable layer, D , from Fig. 1, is $\sim 10 \text{ km}$. Simple reasoning⁹ gives the approximate condition

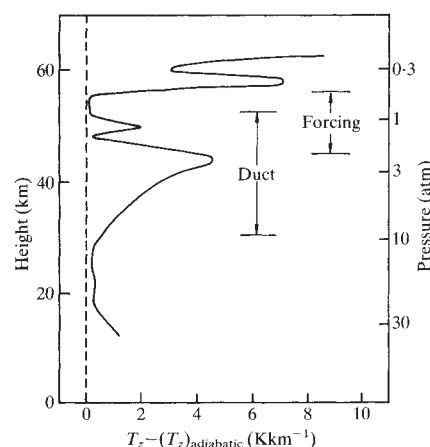


Fig. 2 Temperature gradient profile indicating the stability of the Venus atmosphere, as measured by the Pioneer Venus large probe⁶. Approximate pressures are indicated along the right-hand side. The stable layer at $\sim 40\text{-km}$ elevation acts as a duct for horizontally propagating waves. Forcing within the indicated range was imposed at a point on the equator.

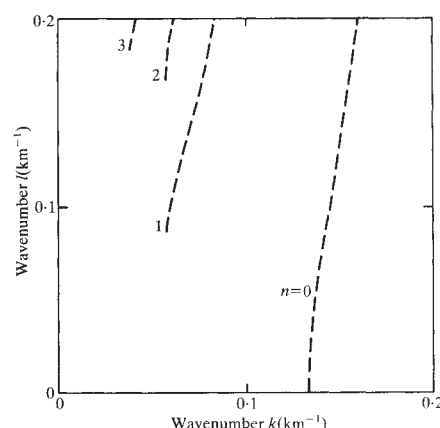


Fig. 3 Locations of resonant or near-resonant response in the east, north wavenumber plane.

for eastward propagating trapped eigenmodes

$$k^2 \approx \frac{N^2}{\bar{u}^2} - \left(n + \frac{1}{2}\right)^2 \frac{\pi^2}{D^2} \quad (2)$$

where n is the number of nodes in the vertical and k is the eastward wavenumber. This relation gives a mode for $n = 0$ with a wavelength of about 30 km . There are no solutions for larger values of n . Thus there is only one mode that can propagate eastward relative to the mean flow. Detailed numerical experiments give the same result, but show a wavelength of $\sim 50 \text{ km}$ for the $n = 0$ mode instead of 30 km .

To perform detailed numerical computations we used the smoothed analytical profiles of $\bar{u}(z)$ and $T(z)$ for Venus published by Young *et al.*¹¹ The equations of motion for linear non-hydrostatic internal gravity waves in a compressible medium whose properties vary only in height were employed. Fourier components of the form $\exp(ikx + ily)$ were studied, where x is eastward and y is northward. A forcing was applied simulating a delta function disturbance at $x = 0, y = 0$. Various forcing distributions in the vertical were used, and the response peaks in the k, l plane were found to be insensitive to this as long as the forcing was within or near the duct. For example, a delta function in height located at the midpoint of the convecting layer is sufficient. To apply boundary conditions in the vertical, we truncated the model atmosphere at two heights well outside the ducting layer and applied a radiation condition.

Fig. 4 Pattern of wave fronts. The fronts with envelope of smaller angle are due to the $n=0$ vertical mode, and the wider one to the $n=1$ vertical mode. Note the caustics at the edges of the wakes.

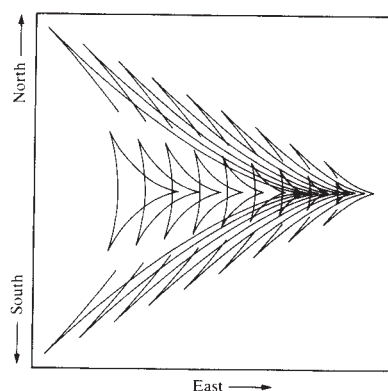


Figure 3 shows curves in the k, l plane where resonance or near resonance is achieved. Consistent with the approximate predictions the $n=1$ and higher modes do not exist for $l=0$. Inspection of the structure of these modes shows that they propagate vertically away from the duct layer for $l=0$ and also for small values of k . Using the lines in Fig. 3, the wave pattern at large distances from the disturbance in the x, y plane can be calculated by standard asymptotic methods¹², and is displayed in Fig. 4 for the modes $n=0$ and $n=1$. The shape of the pattern for the $n=0$ mode is entirely analogous to a ship wake on deep water. The curved phase fronts crossing the line $y=0$ represent the $n=0$ solutions discussed above. They clearly have counterparts in the Venus image of Fig. 1. Another noteworthy feature of the predicted wave patterns are the caustics at the edges of the wakes for both modes. These will be locations of especially large amplitude, and therefore locations where nonlinear effects may be especially important. The general tilt of the spiral pattern on the planet is probably due to the slope of these caustics. As we have argued that the observed pattern is formed of remnants of nonlinear overturning events, we should expect that features of the wave pattern recur at irregular intervals in longitude around the planet. Thus the pattern of Fig. 4 should be imagined to be displaced horizontally and superposed upon itself at intervals. This indeed is what is observed on the planet. Occasionally the arc-shaped waves across the equator are seen to be superposed on the tilted linear features which we have attributed to caustics.

To make an important contribution to the general momentum balance of the mean flow, the wave Reynold's stress due to correlation of eastward and northward velocities, $u'v'$, must be of the same order of magnitude as that due to the mean poleward drift, $\bar{u}\bar{v}$. The mean poleward circulation near the clouds, $\bar{v}$, is presumably due to Hadley cell overturning as a consequence of solar heating at low-latitudes. It is easy to estimate from the solar heating rate and the atmospheric stability that $\bar{v}$ is of the order of 1 m s^{-1} (ref. 13). This is consistent with measurements at the duct height³. Because $\bar{u} \approx 100 \text{ m s}^{-1}$, we find that perturbation velocities u', v' of $\sim 10 \text{ m s}^{-1}$ are required. This is slightly larger than the vertical velocities measured by the balloons, but cannot be ruled out. In addition, there is a question of how intermittent the waves may be. Better data are needed. The fact that the wind speed near the duct is of the same order as the maximum propagation velocity of waves suggests that the requirement for the existence of waves may fix the speed. Unfortunately a purely theoretical approach to the difficult problem of the generation of turbulence by a mountain disturbance and then secondary generation of ducted waves is unlikely to provide a definitive test of these ideas in the near future. Further *in situ* measurements of atmospheric motions, at higher latitudes and at several heights, would be extremely valuable.

3. Counselman, C. C. *et al. J. geophys. Res.* **85**, 8026-8030 (1980).
4. Schubert, G. in *Venus* (eds Hunten, D., Colin, L., Donahue, T. M. & Moroz, V. I.) 681-765 (University of Arizona Press, Tucson, 1983).
5. Knollenberg, R. G. & Hunten, D. M. *J. geophys. Res.* **85**, 8039-8058 (1980).
6. Seiff, A. in *Venus* (eds Hunten, D., Colin, L., Donahue, T. M. & Moroz, V. I.) 215-279 (University of Arizona Press, Tucson, 1983).
7. Covey, C. C. & Schubert, G. *Nature* **290**, 17-20 (1981).
8. Belton, M. J. S. *et al. J. atmos. Sci.* **33**, 1383-1393 (1976).
9. Tolstoy, I. *Wave Propagation* (McGraw Hill, New York, 1973).
10. Schubert, G. & Walterscheid, R. L. *J. atmos. Sci.* **41**, 1202-1213 (1984).
11. Young, R. E. *et al. J. atmos. Sci.* (in the press).
12. Lighthill, J. *Waves in Fluids* (Cambridge University Press, 1978).
13. Gierasch, P. J. *J. atmos. Sci.* **32**, 1038-1044 (1975).

Substitution for copper in a high- T_c superconductor $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$

Yoshiteru Maeno*, Tsukasa Tomita*, Makoto Kyogoku*, Satoshi Awaji*, Yuji Aoki*, Kenichi Hoshino†, Asao Minami† & Toshizo Fujita*

* Department of Physics and † Department of Geology and Mineralogy, Faculty of Science, Hiroshima University, Hiroshima 730, Japan

In spite of an overwhelming amount of research work¹ triggered by the discovery² of superconductivity above 90 K in an Y-Ba-Cu oxide, fundamental questions regarding the origin and mechanism of the high- T_c superconductivity in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ remain to be answered. Specifically, it has not been decided which of the two Cu sites mainly carries superconducting electrons. As an approach to the problem, we investigated superconductivity in a series of alloyed compounds, $\text{YBa}_2(\text{Cu}_{1-x}\text{M}_x)_3\text{O}_{7-\delta}$ with $\text{M} = \text{Fe}, \text{Co}, \text{Ni}, \text{Zn}$ and Ga , in which metals M are partially substituted for Cu. The electrical resistivity measurements, as well as the X-ray analysis at low temperatures, show that T_c is not sensitive to the orthorhombic-tetragonal phase transition induced by the impurities, thus strongly suggesting that superconductivity along the one-dimensional Cu-O chains is unlikely.

The crystal structures of both orthorhombic and tetragonal $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$, including the occupation factors of oxygen, have been precisely determined by neutron diffraction experiments³. There are two distinctive sites for copper in both structures: the first site, Cu-1, is between the two Ba layers and believed to form one-dimensional Cu-1-O chains only in the orthorhombic phase. The second site, Cu-2, is between the Y and Ba layers and constitutes a nearly two-dimensional sheet in both the orthorhombic and tetragonal phases.

We have demonstrated in previous work⁴ that a significant amount of Ni is soluble in the Cu sites to form $\text{YBa}_2(\text{Cu}_{1-x}\text{Ni}_x)_3\text{O}_{7-\delta}$. With $x \leq 0.09$ the crystal structure at room temperature remains orthorhombic and superconductivity is not substantially suppressed by a large amount of disorder introduced by Ni ions. Here, ceramic samples of $\text{YBa}_2(\text{Cu}_{1-x}\text{M}_x)_3\text{O}_{7-\delta}$ with $\text{M} = \text{Fe}, \text{Co}, \text{Ni}, \text{Zn}$ and Ga were prepared by solid-state reaction in air. Starting materials were powders of Y_2O_3 , BaCO_3 and CuO , and of Fe_2O_3 , Co_2O_3 , NiO , ZnO and Ga_2O_3 . The powders were thoroughly mixed, pressed into pellets, and heated at 900°C for 12 h. The pellets were then pulverized, reformed into pellets, heated again at 900°C for 30 h, and cooled to room temperature in the furnace in 24 h. Composition of the crystals with $x = 0.033$ and $\text{M} = \text{Fe}, \text{Co}, \text{Ni}$ and Zn were determined by electron probe microanalysis (EPMA)⁴. The results show that the metals M were indeed substituted for Cu in the crystals with the concentration essentially equal to the nominal concentration.

Figure 1 shows the temperature dependence of the resistivity for an undoped sample ($x=0$), and selected samples with $x=0.033$, measured by the d.c. four-probe method. Above the transition, the undoped sample and the doped samples with $\text{M} = \text{Ni}$ and Zn exhibit metallic behaviour, whereas the samples with $\text{M} = \text{Fe}, \text{Co}$ and Ga show 'semiconductor' behaviour with

Received 30 March; accepted 2 June 1987.

1. Linkin, V. M. *et al. Science* **231**, 1417-1419 (1986).
2. Blamont, J. E. *et al. Science* **231**, 1422-1425 (1986).

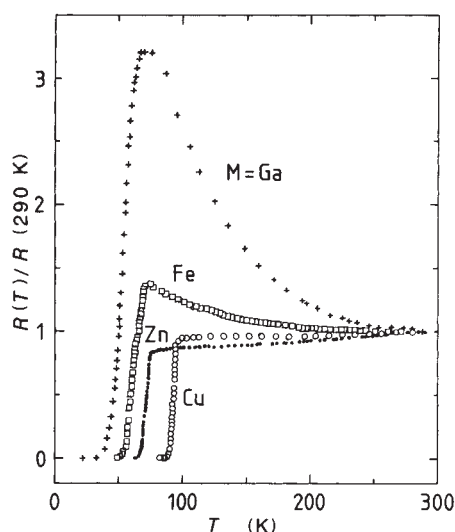


Fig. 1 Temperature dependence of the resistance of $\text{YBa}_2(\text{Cu}_{1-x}\text{M}_x)_3\text{O}_{7-\delta}$ with $x = 0.033$ and $\text{M} = \text{Cu}, \text{Zn}, \text{Fe}$ and Ga .

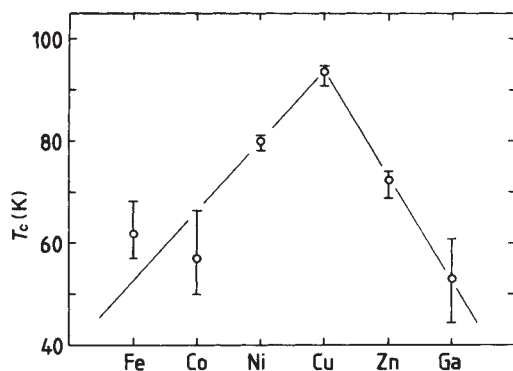


Fig. 2 Superconducting transition temperatures, measured from resistivity, for $\text{YBa}_2(\text{Cu}_{1-x}\text{M}_x)_3\text{O}_{7-\delta}$ with $x = 0.033$ and $\text{M} = \text{Fe}, \text{Co}, \text{Ni}, \text{Cu}, \text{Zn}$ and Ga . The temperatures for 10 and 90% drops in resistivity are indicated by vertical bars.

a negative temperature derivative of resistivity, $d\rho/dT < 0$. The sign of $d\rho/dT$ depends both on heat treatment and impurities, although we presented in this letter the results for samples prepared by identical procedure. In Fig. 2, the superconducting transition temperature for $x = 0.033$ as determined by the resistivity measurements are summarized for various substituted metals with vertical bars indicating the temperatures at which 10 and 90% drops in resistivity occur. The decrease of $T_c(x)$ is observed to be approximately linear in x for small x , with the initial slope $-T_c^{-1}(0)dT_c/dx = 4.4$ and 6.9 , respectively for Ni and Zn, and 10.3 , 11.7 and 13.0 , respectively for Fe, Co and Ga. By comparison, the initial slope for a K_2NiF_4 -type superconductor, $\text{La}_{1.85}\text{Ba}_{0.15}(\text{Cu}_{1-x}\text{M}_x)\text{O}_{4-\delta}$, is 22 and 40 , respectively for $\text{M} = \text{Ni}$ and Zn . These values are about a factor of five larger than those for the corresponding yttrium compounds. The large difference suggests that in the yttrium compounds the substitution with small x does not occur equivalently into both Cu sites, but in favour of one of the Cu sites which is not essential for superconductivity. It is interesting to note that in both the yttrium and lanthanum compounds, doping with non-magnetic Zn has a slightly stronger effect on the suppression of T_c than that with Ni. It should also be noted that magnetic ions Fe and Co have essentially the same effect as a non-magnetic Ga ion on the reduction of T_c .

In addition to the intrinsic impurity effects, the observed reduction in T_c by doping may partly be attributed to various other factors, especially to those associated with oxygen content. Aside from the change in T_c due to the variation in δ , variation in local distribution of oxygen around the impurities is expected

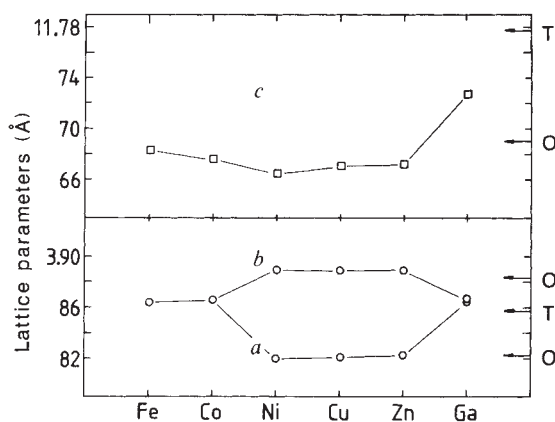


Fig. 3 Lattice parameters for $\text{YBa}_2(\text{Cu}_{1-x}\text{M}_x)_3\text{O}_{7-\delta}$ with $x = 0.033$ and $\text{M} = \text{Fe}, \text{Co}, \text{Ni}, \text{Cu}, \text{Zn}$ and Ga . Reported values (ref. 3) of the lattice parameters for undoped orthorhombic (O) and tetragonal (T) phases are shown by arrows on the right ordinate.

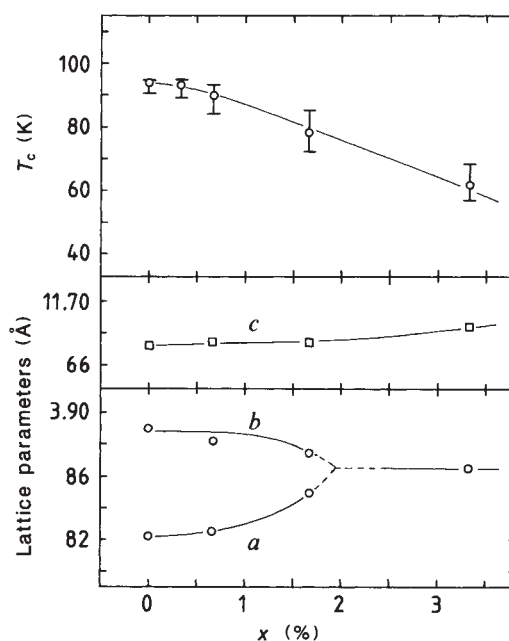


Fig. 4 Concentration dependencies of the superconducting transition temperature T_c and the lattice parameters at room temperature for $\text{YBa}_2(\text{Cu}_{1-x}\text{Fe}_x)_3\text{O}_{7-\delta}$. Doping with Fe induces an orthorhombic-tetragonal phase transition at $x \approx 0.02$. The variation of T_c with x is smooth and insensitive to the structural phase transition.

to change T_c . Although these effects have not been studied in this work, the results presented here are for samples prepared by identical procedures and are valid in representing the effects of doping as a whole.

The X-ray powder diffraction peaks are all indexed for the $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ -type structure. Figure 3 shows the lattice parameters for the undoped and various doped samples, determined at room temperature. For the undoped sample, they are $a = 3.82 \text{ \AA}$, $b = 3.89 \text{ \AA}$ and $c = 11.67 \text{ \AA}$. The lattice parameters for the undoped oxides are indicated by arrows on the vertical axis: they are for an orthorhombic phase with $\delta = 0.31$ and for a tetragonal phase with $\delta = 0.68$. The tetragonal phases induced by doping with Fe, Co and Ga are very different from the tetragonal phase known for an undoped Y-Ba-Cu oxide with a large amount of oxygen deficiency. The c -axis does not stretch so much across the orthorhombic-tetragonal phase transition induced by the Fe-doping of $x = 0.033$, whereas, accompanied by greater oxygen deficiency in the undoped tetragonal phase³, it is larger by $0.09 \text{ \AA}$.

Figure 4 shows T_c , measured from resistivity, and the lattice

parameters at room temperature, plotted against the Fe-concentration x . The critical concentration of Fe for the orthorhombic-tetragonal phase transition is $x_c \approx 0.02$. The X-ray analysis indicates that the structure for $x(\text{Fe}) = 0.017$ remains close to the phase boundary down to 77 K. (For the undoped oxide, the orthorhombic-tetragonal phase transition is known to occur at 943 K in air.) The reduction in T_c with x is not sensitive to the orthorhombic-tetragonal phase transition. One might argue that samples with tetragonal X-ray patterns may contain a small fraction of orthorhombic domains which are responsible for the observed T_c . However, the preliminary measurements indicate that the critical current density for the doped tetragonal samples is comparable to that for the orthorhombic samples, suggesting the bulk nature of superconductivity in the tetragonal phase.

We have shown that $M = \text{Fe, Co, Ni, Zn}$ and Ga can be doped in the Cu sites to constitute a series of alloyed compounds $\text{YBa}_2(\text{Cu}_{1-x}\text{M}_x)_3\text{O}_{7-\delta}$. The reduction in T_c is insensitive to whether the dopant is magnetic or non-magnetic. Rather, the effects of doping with Ni and Zn are distinguished from those with Fe, Co and Ga in various aspects. Doping with a small amount of Fe, Co or Ga induces an orthorhombic-tetragonal phase transition. In contrast, Ni is soluble at least up to the concentration of $x = 0.09$ with no detectable change in the orthorhombic lattice parameters⁴. A possible interpretation is that the valencies of the doped ions (if integer valencies can be assigned at all), most likely 2^+ for Ni and Zn, and 3^+ for Ga and probably 3^+ (or 4^+) also for Fe and Co, or properties related to the valency, such as differences in ionic radii, are more important in determining impurity effects. Isomer shifts in Mössbauer spectra, preliminarily taken on ^{57}Fe -doped samples support this conclusion on the valency of the doped Fe ions.

Based on these observations, we consider one-dimensional superconductivity along Cu-1-O chains to be very unlikely. We speculate that the superconductivity in $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ is essentially of a two-dimensional nature, similar to the situation in $(\text{La}_{1-x}\text{Ba}_x)_2\text{CuO}_{4-\delta}$. To explain a large enhancement of T_c in the yttrium compound, compared with the lanthanum compound, it should be noted, in addition to the regularly ordered frame of Y and Ba, that the oxygen vacancies in the yttrium compound are concentrated in the Cu-1 layer and there are few oxygen vacancies in the Cu-2-O sheets.

Work is under progress to determine the oxygen deficiency δ for various doped samples by chemical and other analyses.

We thank Professor A. Ito at Ochanomizu University for making Mössbauer experiments. This work is supported by Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

Received 18 June; accepted 6 July 1987.

1. Jap. J. appl. Phys. **26**, L311-L525; L601-L881 (1987).
2. Wu, M. K. et al. Phys. Rev. Lett. **58**, 908-910 (1987).
3. Izumi, F. et al. Jap. J. appl. Phys. **26**, L649-L651 (1987).
4. Maeno, Y. et al. Jap. J. appl. Phys. **26**, L774-L776 (1987).
5. Yukino, K. et al. Jap. J. appl. Phys. **26**, L869-L870 (1987).

Melting of peridotite at uppermost lower-mantle conditions

Eiji Ito & Eiichi Takahashi

Institute for Study of the Earth's Interior, Okayama University, Misasa, Tottori-ken 682-02, Japan

An episode of extensive melting early in Earth's history has been postulated on several grounds, but is still controversial¹⁻⁹. To investigate the effects of such a 'magma ocean', we have melted a mantle peridotite at conditions equivalent to those obtaining at the top of the lower mantle (25 GPa, 2,000-3,000 °C). As predic-

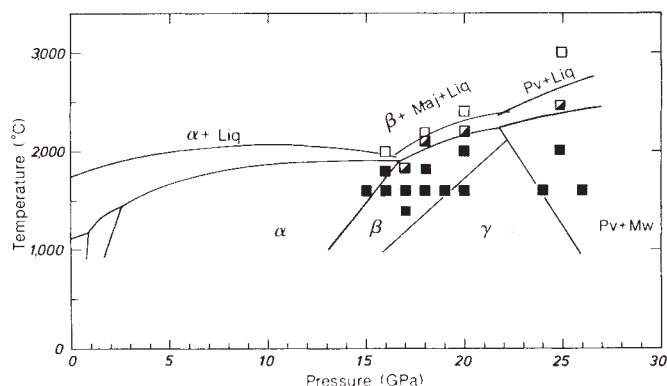


Fig. 1 Melting and sub-solidus phase relations of peridotite PHN-1611 up to 25 GPa. Experimental results below 13 GPa are given in ref. 20; only data points above 15 GPa are plotted. Solid symbols, sub-solidus runs; half-solid symbols, partial melting; open symbols, hyperliquidus runs. Dominant sub-solidus minerals are shown with symbols: α , olivine; β , modified spinel; γ , spinel; Pv, perovskite; Mw, magnesiowüstite. Other abbreviations as in Table 1. See ref. 21 for details of sub-solidus experiments.

ted^{10,11}, MgSiO_3 -rich perovskite is the liquidus phase. A small amount of CaSiO_3 -rich perovskite¹² was observed, and may be the second liquidus phase. Using the Mg/Si, Al/Si, Ca/Si and Ca/Al ratios as constraints, we show that primitive upper-mantle compositions^{1,13-15} can be derived from a chondritic source with a small degree of MgSiO_3 -perovskite fractionation (10-20 wt %).

All of the experiments reported here were made with the USSA-5000 apparatus at Okayama University (see Fig. 1 of ref. 16 for details); experimental techniques have been reported elsewhere^{9,16-21}. The starting material was a sheared garnet lherzolite xenolith from Lesotho kimberlite (PHN-1611)²². Melting phase relations of the peridotite up to 13 GPa were reported in ref. 20; only new experimental data points obtained above 15 GPa are plotted in Fig. 1. Details of the sub-solidus experiments were reported in ref. 21. Both majorite (garnet) and olivine (α -phase) appeared as parallel growths on the melt-crystal interface at 15-16 GPa. At 17 GPa, majorite becomes the first liquidus phase, followed by modified spinel (β -phase). At 20 GPa, a hydrous silicate of approximate composition $(\text{Mg, Fe})_3\text{SiO}_5(\text{OH})_n$ (phase-B)²³ was observed at the melt-crystal interface, along with majorite and modified spinel (Table 1). These results are in agreement with a previous melting study on a chondritic composition up to 20 GPa (ref. 24). The melting experiments in the pressure range 15-20 GPa indicate that the cotectic boundary between the orthosilicates (α - and β -phase) and metasilicates (majorite) lies very close to the bulk chemistry of the peridotite. A small amount of water was found to be present in the experimental assembly, even after it was carefully dried by a torch.

At 25 GPa and below the solidus (1,600-2,000 °C) the peridotite consists of MgSiO_3 -perovskite, magnesiowüstite, majorite and a small amount of an unquenchable Ca-rich phase (either diopside^{18,21} or CaSiO_3 ²⁵). At ~2,500 °C and 25 GPa, melt and solid are almost completely separated (Fig. 2a); see also Fig. 7 of ref. 9). The higher-temperature portion is occupied by quenched liquid, consisting of a very fine aggregate of perovskite, magnesiowüstite and glass. The lower-temperature portion is occupied by large MgSiO_3 -perovskite crystals. Small blebs (5-30 μm diameter) of a CaSiO_3 -rich phase¹² are scattered in the large MgSiO_3 -perovskite crystals >150 μm away from the melt-solid interface (Fig. 2b).

The compositions of the MgSiO_3 -perovskite, CaSiO_3 -phase and quenched liquid were determined by electron-probe microanalysis (Table 1). Because small (<1 μm) magnesiowüstite grains were lost from the quenched liquid during grinding and polishing (see Fig. 2 legend), the analyses of the quenched liquid are low in total (70-80 wt %) and probably deficient in MgO and FeO. The average composition of the liquid, normalized to

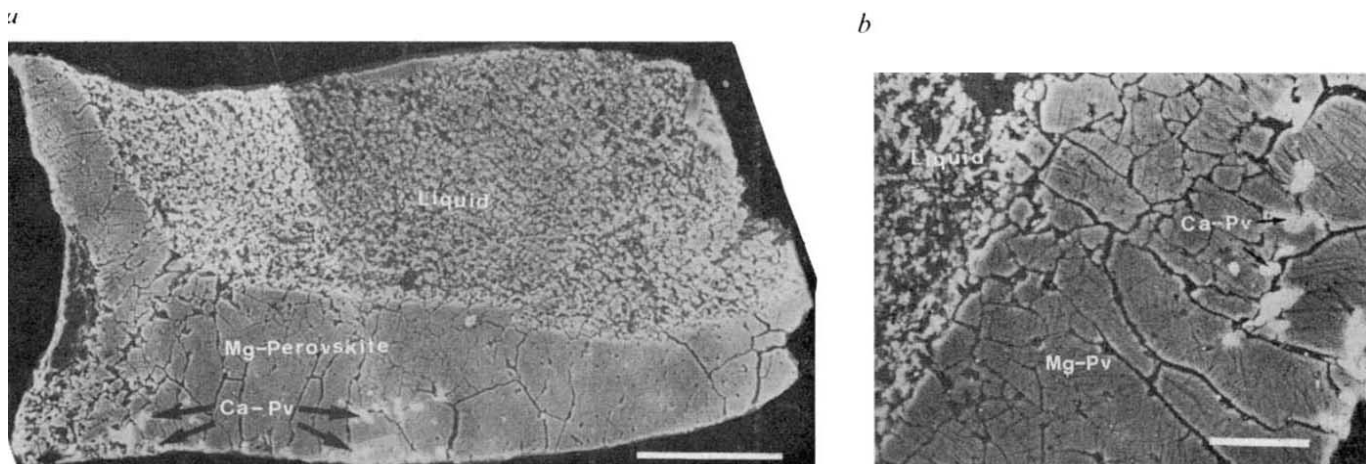


Fig. 2 Back-scattered electron images of the experimental charge quenched at 25 GPa and 2,500 °C. Scale bars: *a*, 300 μm ; *b*, 50 μm . Partial melting and separation of the liquid (higher-temperature) and solid (lower-temperature) regions were completed within the run duration of 10 min. The solid residue consists almost exclusively of large MgSiO_3 -perovskite crystals. In the lowest-temperature region, >150 μm away from the melt-solid interface, small grains of a CaSiO_3 -rich phase (glass) with an unquenchable cubic perovskite structure¹³ are observed. The quenched liquid is very friable because of the presence of a large amount of magnesiowüstite (small equant grains, <1 μm). The electron-microprobe analysis of the liquid (Table 1) may be depleted in MgO and FeO owing to the depletion of magnesiowüstite from the polished surface.

100%, is listed in Table 1, together with the estimated liquid composition derived from a mass-balance calculation (assuming 40 wt % MgSiO_3 -perovskite and 60 wt % liquid). Both liquid and solids are enriched in lanthanum and chromium, owing to contamination from the LaCrO_3 heater, through the sintered diamond sample container.

The results of the melting experiments at 25 GPa may be summarized as follows: MgSiO_3 -perovskite is the principal liquidus phase in the peridotite composition; a small amount of CaSiO_3 -perovskite may crystallize as the second liquidus phase; and at small degrees of partial melting near the solidus, a 'super-ultramafic' magma (enriched in $\text{MgO} + \text{FeO}$ relative to $(\text{Mg}, \text{Fe})_2\text{SiO}_4$) may be produced by eutectic melting of MgSiO_3 -perovskite and magnesiowüstite. In a similar experiment using an $(\text{Mg}_{0.8}\text{Fe}_{0.2})_2\text{SiO}_4$ starting material at 25 GPa, MgSiO_3 -rich perovskite appeared as the first liquidus phase.

It is commonly assumed that the bulk Earth composition is chondritic^{1-3,5}. Compared with chondrite types C1, L, H and E,

however, upper-mantle peridotites (UMPs) have systematically higher Mg/Si ratios (Fig. 3*a,b*). Consideration of the depletion factors of elements in UMP (pyrolite) compared with C1 chondrite as a function of volatility led Ringwood^{1,26} to explain the higher Mg/Si ratios in UMPs by postulating volatilization of silicon from the Earth before or during accretion. An alternative explanation is that the mantle is chemically stratified (ultramafic upper mantle and silica-rich lower mantle) as a result of majorite and/or perovskite fractionation in the primordial terrestrial magma ocean^{5,7-9,15,24,27}. The present experiments allow us to test the latter hypothesis.

Figure 3 summarizes the major-element ratios (Mg/Si, Ca/Si, Al/Si and Ca/Al) of upper-mantle peridotites (shaded area)^{2,3,13,15,28}, estimated primitive upper-mantle compositions (MA²⁸, R¹, J¹³, PN¹⁵) and chondrites (types E, L, H and C1)^{3,15}. Majorite (open stars) and MgSiO_3 -perovskite (solid stars) which are the primary liquidus phases at 16–20 GPa and >24 GPa, respectively, are also plotted in Fig. 3. Because of the high Al/Si

Table 1 Chemical compositions of run products and estimated bulk silicate Earth composition

	Starting composition (PHN-1611)	25 GPa, ~2,500 °C								Estimated BSE composition			
		20 GPa, 2,200 °C											
		Maj	B	β	Liq	Mg-Pv (Avg. of 8)	Ca-Pv (Avg. of 7)	Liq. (Avg. of 3)	(Mass balance)§	BSE-M1 (This work)	R ¹	M&A ³⁰	Z&H ³
SiO_2	44.54	51.4	32.4	40.2	42.6	$54.6 \pm 0.8^\dagger$	44.3 ± 0.1	40.9 ± 0.3	(37.8)	46.7	45.1	47.9	46.0
TiO_2	0.25	0.2	0.0	0.1	0.5	0.25 ± 0.2	0.17 ± 0.01	0.21 ± 0.02	(0.3)	0.23	0.2	0.20	0.18
Al_2O_3	2.80	9.5	0.7	1.5	1.6	3.2 ± 0.2	3.2 ± 0.3	3.3 ± 0.1	(2.5)	3.9	3.3	3.9	4.0
FeO^*	10.24	4.4	8.5	8.5	15.0	3.6 ± 0.6	0.3 ± 0.2	10.2 ± 0.2	(14.7)	7.0	8.0	8.9	7.6
MnO	0.13	0.1	0.1	0.1	0.2	—	—	—	—	0.13	0.15	0.14	0.13
MgO	37.94	31.0	58.9	48.1	33.2	36.5 ± 1.0	1.1 ± 0.4	31.9 ± 0.6	(38.9)	38.2	38.1	34.1	37.8
CaO	3.32	2.3	0.0	0.0	5.7	0.42 ± 0.04	41.6 ± 1.2	8.9 ± 0.3	(5.3)	3.1	3.1	3.2	3.2
Na_2O	0.34	0.0	0.1	0.0	0.7	0.03 ± 0.01	0.05 ± 0.01	0.12 ± 0.01	(0.5)	0.29	0.4	0.25	0.33
K_2O	0.14	—	—	—	—	—	—	—	—	0.03	0.03	0.03	0.03
Cr_2O_3	0.29	1.2	0.6	0.5	0.5	1.4 ± 0.3	0.8 ± 0.2	2.7 ± 0.1	—	0.5	0.4	0.9	0.5
La_2O_3	—	—	—	—	—	0.00	8.5 ± 1.4	1.8 ± 0.1	—	—	—	—	—
Total	99.99	100.1	101.3	99.0	100.0	100.0	100.0	100.0	(100.0)	100.0	98.8	99.5	99.8
Atomic Mg/(Mg + Fe)	0.868	0.925	0.925	0.910	0.798	0.964	—	0.847	(0.825)	0.91	0.89	0.87	0.90
$K_D(\text{Fe}/\text{Mg})^\ddagger$	—	0.31	0.32	0.39	—	$0.31(0.26)^\S$	—	—	—	—	—	—	—
Al/Si§	0.071	0.208	—	0.042	0.043	0.066	0.080	0.091	(0.075)	0.094	0.083	0.092	0.100
Mg/Si	1.099	0.799	2.347	1.545	1.006	0.888	—	1.007	(1.329)	1.057	1.091	0.919	1.061
Ca/Si	0.114	0.067	—	—	0.205	0.012	1.434	0.333	(0.215)	0.100	0.105	0.102	0.107
Ca/Al	1.603	0.322	—	—	4.815	0.177	17.85	3.645	(2.865)	1.065	1.270	1.109	1.069

Maj, majorite; B, phase-B²³; B, modified spinel; liq, liquid; Pv, perovskite.

* Total iron as FeO. † Errors listed are ± 1 standard deviation. ‡ $K_D(\text{Fe}/\text{Mg}) = (\text{Fe}/\text{Mg})^{\text{crystal}}/(\text{Fe}/\text{Mg})^{\text{liquid}}$. § Elemental ratios are given as weight ratios of the metals.

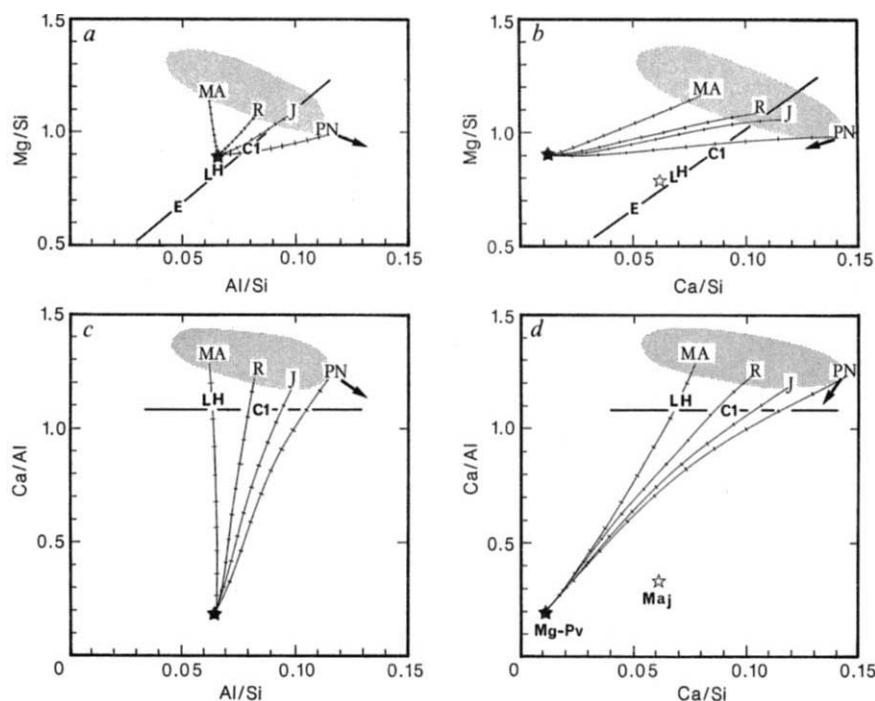


Fig. 3 Major-element ratios (metal weight ratios) of upper-mantle peridotites (shaded area) and chondrites (types E, H, L and C1). See ref. 3 for similar diagrams. Primitive mantle compositions estimated by Maaløe and Aoki²⁸ (MA), Ringwood¹ (R) Jagoutz *et al.*¹³ (J) and Palme and Nickel¹⁵ (PN) are also shown. Stars are the compositions of the MgSiO₃-perovskite (solid stars) and majorite (open stars) that are on the peridotite liquidus at 25 and 20 GPa, respectively (Table 1). Because of the high Al/Si ratio of the majorite, it cannot be shown in *a* and *c*. To deduce the identity of phases that could be fractionated from the primordial terrestrial magma ocean to yield ultramafic upper-mantle peridotite, we performed binary mixing calculations between the proposed primitive mantle compositions (MA, R, J and PN) and MgSiO₃-perovskite (or majorite). Mixing trends between primitive mantle and MgSiO₃-perovskite are shown with tie lines; the directions of the mixing trends between composition PN and majorite are shown with arrows. A mixture of 85 wt % composition J and 15 wt % MgSiO₃-perovskite satisfies the cosmochemical constraints in all of the diagrams.

ratio of the majorite (0.208; see Table 1), it is not plotted in Fig. 3a, c. Calculations of binary mixing between majorite or MgSiO₃-perovskite and primitive upper-mantle compositions (MA, R, J, PN) were carried out; only the results for MgSiO₃-perovskite are shown in Fig. 3. Trends of majorite addition are indicated with arrows for the estimated primitive mantle composition PN.

Calcium and aluminium are the most refractory of the major elements; the significant deviation in Ca/Al ratio of UMPs (1.2–1.4) from chondrites (1.13–1.08) (Fig. 3c, d) is therefore difficult to explain by selective volatilization¹⁵. Palme and Nickel¹⁵ and Ohtani *et al.*²⁴ explained this deviation by majorite fractionation in the primordial magma ocean, but Fig. 3a, c shows that binary mixing of UMPs with majorite cannot yield a chondritic Al/Si ratio at given Mg/Si and Ca/Al ratios. Thus, majorite fractionation is not compatible with the hypothesis that the bulk silicate Earth (BSE) has chondritic element abundances. The same conclusion has been obtained by Kato *et al.*²⁹, who also determined partition coefficients of rare-earth elements between majorite and ultramafic magma at 16 GPa and 2,000 °C. According to their partition coefficients²⁹, UMPs should exhibit light-rare-earth enrichment and heavy-rare-earth depletion if extensive majorite fractionation had occurred in the terrestrial magma ocean. Because primitive UMPs show unfractionated rare-earth patterns, with 1.5–3 times chondritic abundances^{2,23,14} (when normalized by Mg in C1 chondrite, the rare-earth abundance in primitive UMP is 1.16 times C1)¹³, majorite fractionation in the magma ocean cannot have had a significant role in the derivation of UMPs²⁹.

Because MgSiO₃ perovskite has a very low Ca content (Table 1), fractionation of this phase from the magma ocean will rapidly increase the Ca/Al ratio of the liquid (Fig. 3c, d). Fractionation of 10–20 wt % of MgSiO₃-perovskite can account for the observed deviation in Ca/Al of UMPs from chondrites. Although binary mixing of 80 wt % pyrolite (composition R)¹ and 20 wt % MgSiO₃-perovskite gives Ca/Al, Al/Si and Ca/Si ratios approximately equal to those of C1 chondrite (Fig. 3c, d), the Mg/Si ratio is significantly higher than that of C1 (Fig. 3a, b). On the other hand, a mixture of 50 wt % PN¹⁵ and 50 wt % MgSiO₃-perovskite is similar to C1 in Mg/Si, Al/Si and Ca/Si but is significantly lower in Ca/Al. Thus, no supposed primitive UMP composition can be derived from the C1 composition by perovskite fractionation alone.

Note, however, that a mixture of 90–80 wt % composition J (ref. 13) and 10–20 wt % MgSiO₃-perovskite lies approximately on the cosmochemical array (Fig. 3). The inferred BSE composition (BSE-M1) is shown in Table 1. Compared with C1, BSE-M1 is slightly higher in Mg/Si, Ca/Si and Al/Si, possibly due to volatilization of silicon. Because of the small distribution coefficient for (Fe/Mg) between MgSiO₃-perovskite and liquid (0.26–0.30; Table 1), our BSE estimate is higher in Mg/(Mg + Fe) than those of previous workers^{1,3,30}.

In our experiment at 25 GPa and 2,500 °C (Fig. 2a), the lanthanum partition coefficient between MgSiO₃-perovskite and liquid was found to be <0.1 (Table 1), in agreement with the estimate by Kato *et al.*³⁰. MgSiO₃-perovskite fractionation will, therefore, concentrate rare-earth elements in the liquid, unless CaSiO₃-perovskite (which accommodates rare-earths as major elements; see Table 1 and ref. 30) appears. Thus, 15 wt % of MgSiO₃-perovskite fractionation would be consistent with the unfractionated and yet higher rare-earth abundances (1.16 times C1) in UMPs compared with chondrites¹³.

We have shown here that, if the primordial terrestrial magma ocean extended down to the lower mantle (>24 GPa), MgSiO₃-perovskite would have been the residual phase. The major-element character of UMPs (including Ca/Al and rare-earth abundances) can be derived from a chondritic source material (BSE-M1) by a small degree of MgSiO₃-perovskite fractionation. It is not yet known whether the Earth experienced an extensive melting stage, but the present work shows that, even if there was a deep magma ocean, the upper mantle may not differ greatly in composition from the lower mantle.

Received 3 March; accepted 22 May 1987.

1. Ringwood, A. E. *Origin of the Earth and Moon* (Springer, New York, 1979).
2. Basaltic Volcanism Study Project. *Basaltic Volcanism on the Terrestrial Planets* (Pergamon, New York, 1981).
3. Zindler, A. & Hart, S. R. *A. Rev. Earth planet. Sci.* **14**, 493–571 (1986).
4. Anderson, D. *J. geophys. Res.* **84**, 6297–6298 (1979).
5. Anderson, D. & Bass, J. D. *Nature* **320**, 321–328 (1986).
6. Nisbet, E. G. & Walker, D. *Earth planet. Sci. Lett.* **60**, 103–113 (1982).
7. Ohtani, E. *Phys. Earth planet. Inter.* **38**, 70–80 (1985).
8. Herzberg, C. T. & O'Hara, M. J. *Geophys. Res. Lett.* **12**, 541–544 (1985).
9. Takahashi, E. *J. geophys. Res.* **91**, 9367–9382 (1986).
10. Kumazawa, M. (abstr.) *Intern. Geodynamics Conf. Tokyo*, 276–277 (1978).
11. Ohtani, E., *Phys. Earth planet. Inter.* **33**, 12–25 (1983).
12. Liu, L. & Ringwood, A. E. *Earth planet. Sci. Lett.* **28**, 209–211 (1975).
13. Jagoutz, E. *et al. Proc. lunar planet. Sci. Conf.* **10**, 2031–2050 (1979).
14. Sun, S. S. *Geochim. cosmochim. Acta* **46**, 179–192 (1982).
15. Palme, H. & Nickel, K. G. *Geochim. cosmochim. Acta* **49**, 2123–2132 (1985).
16. Ito, E., Takahashi, E. & Matsui, Y. *Earth planet. Sci. Lett.* **67**, 238–248 (1984).

17. Ito, E. & Yamada, H. in *High-Pressure Research in Geophysics* (ed. Akimoto, S.) 405-419 (Reidel, Dordrecht, 1982).
18. Ito, E. & Takahashi, E. in *High Pressure Research Applications in Geophysics and Geochemistry* (ed. Manghnani, M. H.) (American Geophysical Union, Washington, 1987).
19. Takahashi, E. & Scarfe, C. M. *Nature* **315**, 566-568 (1985).
20. Scarfe, C. M. & Takahashi, E. *Nature* **322**, 354-356 (1986).
21. Takahashi, E. & Ito, E. in *High Pressure Research Applications in Geophysics and Geochemistry* (ed. Manghnani, M. H.) (American Geophysical Union, Washington, 1987).
22. Nixon, P. H. & Boyd, F. R. in *Lesotho Kimberlites* (ed. Nixon, P. H.) (Lesotho National Development Corporation, Maseru, 1973).
23. Kato, T. & Kumazawa, M. *Geophys. Res. Lett.* **13**, 181-184 (1986).
24. Ohtani, E., Kato, T. & Sawamoto, H. *Nature* **322**, 352-353 (1986).
25. Irifune, T. & Ringwood, A. E. in *High Pressure Research Applications in Geophysics and Geochemistry* (ed. Manghnani, M. H.) (American Geophysical Union, Washington, 1987).
26. Ringwood, A. E. & Kesson, S. E. *The Moon* **16**, 425-464 (1977).
27. Walker, D. *Contr. Miner. Petrol.* **92**, 303-307 (1986).
28. Maaløe, S. & Aoki, K. *Contr. Miner. Petrol.* **63**, 161-173 (1977).
29. Kato, T., Irifune, T. & Ringwood, A. E. *Geophys. Res. Lett.* **14**, 546-549 (1987).
30. Kato, T., Irifune, T. & Ringwood, A. E. *Lunar Sci. Conf. Abstr.* **18**, 483-484 (1987).
31. Morgan, J. W. & Anders, E. *Proc. natn. Acad. Sci. U.S.A.* **77**, 6973-6977 (1980).

The Minoan eruption of Santorini in Greece dated to 1645 BC?

C. U. Hammer*, H. B. Clausen*, W. L. Friedrich† & H. Tauber‡

* Geophysical Institute, University of Copenhagen, Haraldsgade 6, DK-2200 Copenhagen N, Denmark

† Geological Institute, Aarhus University, C. F. Møllersalle 120, DK-8000 Aarhus C, Denmark

‡ Radiocarbon Dating Laboratory, National Museum, Ny Vestergade 10, DK-1471 Copenhagen K, Denmark

The eruption on Santorini (Thera: 36.40° N, 25.40° E) in the Aegean Sea during Late Minoan time is considered the most violent volcanic event in the Mediterranean in the second millennium BC. The eruption buried a number of developing Bronze Age settlements on the island (one of which is presently being excavated at the village Akrotiri) and spread huge amounts of tephra over the eastern Mediterranean and adjacent lands²⁻⁷. This event is therefore an important time marker both for archaeologists and Earth scientists. A dating of the eruption has previously been attempted by archaeological inference^{1,8} and by radiocarbon dating, but the two methods have tended to give ages that deviate by up to 150 years. Here we present a new ice-core dating of the eruption, which suggests an age of 1645 BC based on variations in acid fallout in the annual ice layers in a core drilled at the site Dye 3 (65.18° N, 43.49° W) in South Greenland.

It is generally agreed among volcanologists⁹⁻¹¹ that the Minoan eruption consisted of three or four phases that changed an originally almost circular island called Stronghyle into three ring-islands, Thera, Therasia and Aspronisi, with a huge caldera in the central part (see Fig. 1). Today this caldera has an area of 84.9 km² and a depth of ~300 m below sea-level. The eruption started with a strong Plinian phase producing up to a 7-m-thick layer of air-fall pumice on Santorini. In this initial phase enormous amounts of tephra, dust and gases were ejected into the atmosphere. Wilson¹² estimates that the eruption column was at a height of ~29 km, and thus penetrated well into the stratosphere. The following phases of the eruption also contributed in producing dust and gases, but not in such amounts, nor to such a height as in the Plinian phase. Estimates of the total amount of pyroclastics produced during the eruption vary from 18 to 40 km³ (refs. 10, 13). The Minoan eruption has been compared with the AD 1883 Krakatoa (6.10° S, 105.42° E) and the AD 1815 Tambora (8.25° S, 118.00° E) eruptions, which were two of the strongest volcanic eruptions during the past two centuries. The volcanic explosivity index (VEI) of these eruptions has been given as 6 and 7, respectively¹⁴. The VEI magnitude of the Minoan eruption has been estimated as 6 on the same scale¹⁴. Eruptions of this magnitude are expected to show up in acidity records in the Greenland ice cores. Based on correlations

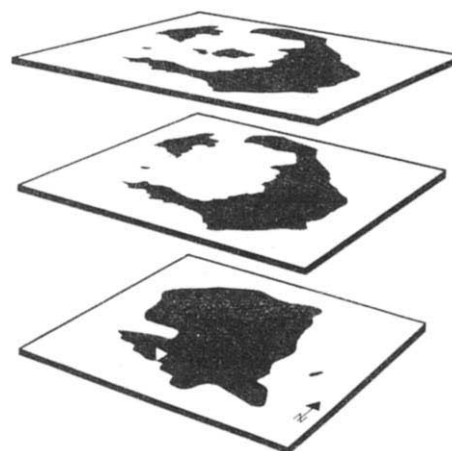


Fig. 1 Effects of the Minoan eruption on Santorini. The three stages show the change of the pre-eruption Stronghyle (the rounded) island (lowest) to the post-eruption shape (middle) and the present shape with the two Kameni islands in the middle of the caldera (upper). Reconstruction of the Stronghyle island after ref. 10. The white triangle (lower) shows the Bronze Age settlement close to present village Akrotiri.

between cultural elements in the destruction level on Santorini and findings from Crete, from the Greek mainland, and from the Cyclades, archaeologists have, up till now, agreed on an age for this catastrophe at ~1500 BC^{1,8}.

In a recent paper, Betancourt¹⁵ has tried to revise the archaeological estimate of the age of the Late Minoan IA period (LMIA), at the end of which the eruption took place, such that the age becomes ~150 years older in agreement with the radiocarbon evidence. Radiocarbon dating of the Minoan eruption has been attempted on a large number of samples connected with the destruction level at the site Akrotiri on the island Thera. The most reliable results are obtained if only dates of short-lived materials (twigs of shrubs and crops harvested shortly before the eruption) directly connected with this destruction level are considered. Sixteen radiocarbon dates of this kind have been published¹⁶⁻¹⁹. They give a statistically consistent pattern with only three outliers (P-1888, P-2560 and P-2561) which deviate by 3.6, 8.3 and 9.7 times the standard deviation, respectively, from the mean of the remaining results. If these outliers are left out, 13 radiocarbon dates of short-lived materials remain. These are listed in Table 1 together with a new date (K-4255) made

Table 1 Carbon-14 dates of short-lived samples associated with the Minoan eruption

Sample no.	Conventional C-14 age BC	Calibrated age BC	Calibrated age $\pm 1\sigma$ BC
P-1885	1296 $\pm$ 48	1520	1610-1460
P-1889	1348 $\pm$ 52	1610-1540	1675-1520
P-1892	1381 $\pm$ 52	1630	1690-1525
P-1894	1357 $\pm$ 65	1615	1680-1520
P-1895	1374 $\pm$ 51	1625	1685-1525
P-2559	1430 $\pm$ 70	1685	1750-1610
P-2565	1370 $\pm$ 60	1620	1685-1520
P-2791	1390 $\pm$ 60	1645	1735-1525
P-2792	1720 $\pm$ 180	2110-2040	2330-1780
P-2793	1350 $\pm$ 140	1610-1540	1750-1430
P-2794	1230 $\pm$ 50	1445	1515-1420
P-2795	1430 $\pm$ 170	1685	1890-1515
K-3228	1390 $\pm$ 55	1645	1735-1525
K-4255	1430 $\pm$ 60	1685	1750-1620
Weighted mean	1359 $\pm$ 17	1615	1630-1530

Calibration according to Pearson and Stuiver²⁰.

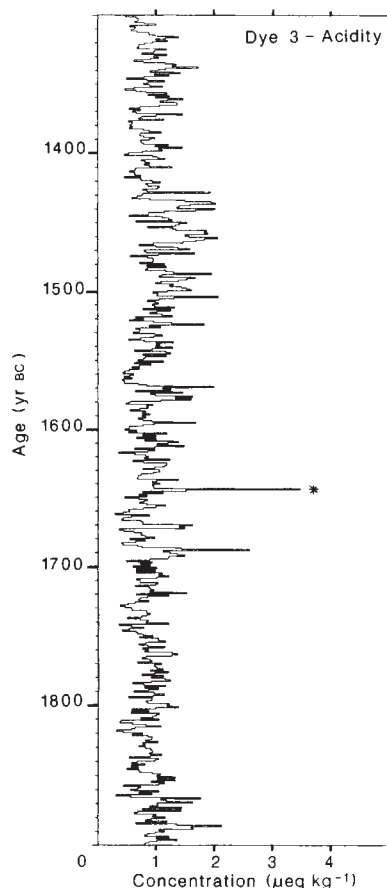


Fig. 2 The annual average H^+ concentration (in $\mu\text{eq kg}^{-1}$ of ice) in the Dye 3 core over the time period 1300–1900 BC; depth interval 1145.49–1280.47 m. The acidity values are based on a continuous record (a few mm resolution) obtained by a special electric conductometric method ECM²⁴ and calibrated accordingly. For the above section of the Dye 3 deep core this calibration procedure tends to underestimate the actual acidities; however, this has only little influence on the general appearance of the acidity profile and mainly affects peaks exceeding $3 \mu\text{eq kg}^{-1}$ (see also Fig. 3 legend). The peak marked by an asterisk is the most likely 'candidate' for the Minoan eruption.

on a twig of *Tamarix* (10 years rings) found in the destruction level at Akrotiri (house 3, structure delta 1). The weighted ($p = 1/\sigma^2$) mean of these dates is 1359 radiocarbon years BC with a standard deviation of ± 17 years, if calculated on the basis of the individual sigma values, or ± 18 years if calculated as a root-mean-square error. If this date is calibrated in accordance with the globally valid high-precision curves of Pearson and Stuiver²⁰, a 1σ band around the weighted mean corresponds to the time interval 1630–1530 BC in calendar years, and a 2σ band corresponds to the interval 1675–1525 BC. The wide time range and the asymmetric distribution around the mean value are due to the irregular course of the calibration curve in this period. We note that due to the form of the calibration curve the inclusion of the younger outlier (P-1888) would not have changed the upper limits of the 1σ and 2σ bands, and that the statistical consistency suggests that a possible influence on the dates from volcanic emanations²¹ is negligible.

The Dye 3 deep ice core presents a complete record of annual layers of snow deposits which have been dated stratigraphically by means of the seasonal variations within each annual layer²². Three different, seasonally varying parameters (isotopic composition, dust content and acidity) were used in order to obtain the most accurate dating possible. Note that the Dye 3 area receives sufficient amounts of annual precipitation to rule out

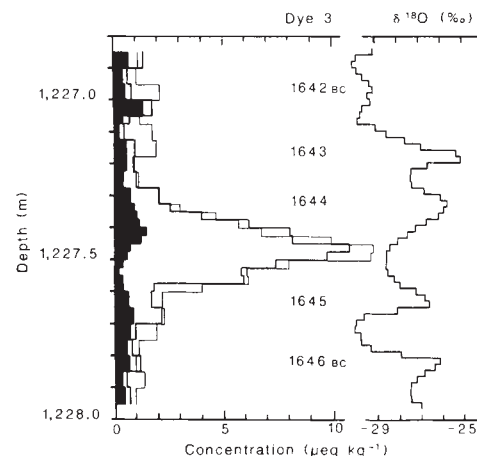


Fig. 3 The $\delta^{18}\text{O}$ curve to the right illustrates the annual ice layers from 1646 BC to 1642 BC as reflected by the seasonal variations in $\delta^{18}\text{O}$. The curves to the left indicate concentrations of NO_3^- (shaded area), of $\text{NO}_3^- + \text{SO}_4^{2-}$ excess (thick line), and of acidity (thin line). The difference between the area to the left of the thick line and the shaded area thus shows concentrations of excess sulphate (that is, sulphate concentration corrected for the sea salt component). The ECM acidity profile was specially calibrated over this ice-core sequence, by pH measurements, to judge accurately the chemical characteristics of the sulphate peak. The NO_3^- (and Cl^- ; not shown) concentrations are not enhanced during the event. The enhanced signal is therefore due to sulphuric acid and must have been caused by a volcanic eruption of considerable magnitude.

the possibility of missing annual snow deposits. A continuous high-resolution acidity record of the entire Dye 3 core, which extends far back into the glacial period, was measured at the site by means of a special electric conductivity technique^{23,24}. The acidity record in the Dye 3 core during the time interval 1300–1900 BC, a period within which the Minoan eruption must have taken place, is shown in Fig. 2.

From a technical point of view this record is almost ideal, but due to the southerly location of the drill site, melt layers associated with warm summers complicate the interpretation of the acidity record. Such melt layers may be associated with acidity peaks particularly enriched in nitric acid, whereas volcanic acidity peaks are mainly composed of sulphuric acid (in some cases together with hydrochloric acid and hydrofluoric acid). A separation between these two types of acidity signals is therefore necessary before the record can be safely interpreted.

Further, due to the general poleward transport in the high-latitude stratosphere²⁵, some acidity signals from high northern eruptions may be absent in ice cores from south Greenland, whereas they are clearly seen in ice cores from central and north Greenland: The AD 1912 Katmai (Novarupta, 58.27°N , 155.16°W) eruption (VEI 6) in the Aleutian Islands thus left clear high-acid signals in north and central Greenland ice cores^{26,27}, but was barely traceable in south Greenland, even though the stratospheric dust veil from this eruption was observed as far south as 34°N . But acids injected into the stratosphere from major volcanic eruptions at intermediate and low northern latitudes are carried by the stratospheric circulation to higher latitudes and therefore should leave a clear acidity signal in ice cores both in south and north Greenland.

Based on acid deposits in the north Greenland Camp Century ice core, a date of 1390 BC has previously been suggested for the Minoan eruption²⁶. However, no acidity peak is present in the Dye 3 ice core at ~ 1390 BC. The 1390 BC acidity signal in the Camp Century core therefore can hardly come from a major mid-latitude eruption, but most likely was caused by a high-latitude eruption like that of Katmai. Moreover, the Camp Century acidity record, was later found to be incomplete, in

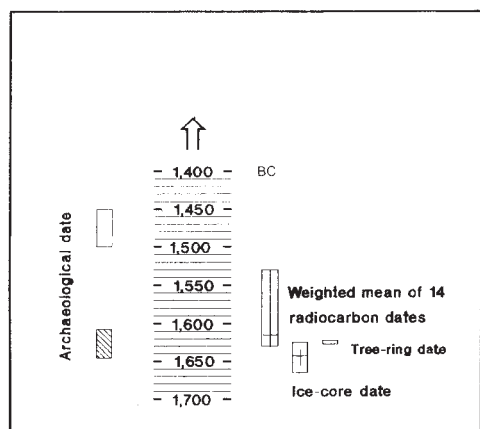


Fig. 4 Dates of the Minoan eruption in calendar years BC obtained by different methods. The open block to the left indicates the traditional archaeological estimate, and the hatched block the revised estimate by Betancourt¹⁵.

particular during the period 1600–2000 BC²⁷, when the eruption is now thought to have occurred.

The record shown in Fig. 2 indicates a number of acidity peaks superimposed on a general background of acidity in the ice layers. The background oscillates between periods of low general acidity (for example, 1553–1568 BC) and periods with enhanced general acidity (for example, 1431–1441 BC). The enhanced acidity background is largely determined by increased summer melting due to the high summer temperature in south Greenland. The individual acidity peaks superimposed on the background are caused either by acid fallout of volcanic origin or by intense melting during the summer months.

Three methods were used to distinguish between these two types of acidity peaks: (1) visual inspection of the core for distinct melt features, (2) investigation of the annual acidity profiles to separate melt layer acidity (usually irregularly distributed over the year) from volcanic acidity (normally high acidity over the entire annual ice layer) and (3) direct chemical analysis (ion chromatography) of the high-acidity signals. The third method is by far the most powerful and leads to an unambiguous separation, but it is very time consuming. Hence, it was only applied to the three acidity signals, which were considered to be the most likely candidates to be caused by a single and violent volcanic eruption: the peaks at 1428 BC, at 1644 BC and at 1688 BC. The peak at 1428 BC does not represent a particularly high acidity, but it is superimposed on a low-acidity background; the peak at 1644 BC is the highest acidity signal and shows a consistent high acidity over more than a year; the 1688 BC peak is the second highest signal, but is superimposed on a high-acidity background. The results of the chemical analysis of these three events show that the 1644 BC signal is the only one with a high content of sulphuric acid and thus of volcanic origin. The strong increase in the sulphate concentration at 1644 BC is shown in Fig. 3. As the amount of sulphuric acid has already increased at the end of the previous year, the actual eruption probably took place during that year, that is, in 1645 BC. The peaks at 1428 BC and 1688 BC were both dominated by nitric acid. We therefore suggest that the 1644 BC acidity peak is due to the Minoan eruption, and that the date of the eruption is 1645 BC with an estimated standard deviation of ± 7 yr, and an estimated error limit of ± 20 yr. This suggestion could be further supported, if a corresponding acidity peak of the same age is found in new deep ice cores from central Greenland. In such an ice core major Alaskan explosive eruptions during 1300–1900 BC should also be clearly represented²⁸. Dates of the Minoan eruption obtained by different methods, including a tentative date derived from analyses of frost damages in tree rings²⁹, are summarized in Fig. 4.

The ice-core analyses were supported by the Commission of European Communities and the Commission for Scientific Research in Greenland. Field work on Santorini (W.L.F.) has been supported by the Carlsberg Foundation. We thank Hans Pichler and Tom Simkin for helpful comments concerning the Minoan eruption. Christos Doumas and Charalambos Sigalaeas are acknowledged for providing radiocarbon samples from Akrotiri and Peter Wagner for determining the Tamarix twig.

Received 20 February; accepted 1 June 1987.

1. Doumas, C. G. *Thera, Pompeii of the Ancient Aegean* (Thames and Hudson, London, 1983).
2. Pichler, H. & Schiering, W. *Nature* **267**, 819–822 (1977).
3. Watkins, N. D. *et al.* *Nature* **271**, 122–126 (1978).
4. Doumas, C. & Papazoglou, D. *Nature* **287**, 322–324 (1980).
5. Keller, J. in *Tephra Studies* (eds Sparks, R. S. J. & Self, S.) 227–244 (Reidel, Dordrecht, 1981).
6. Vinci, A. *Mar. Geol.* **64**, 143–155 (1985).
7. Stanley, D. J. & Sheng, H. *Nature* **309**, 733–735 (1986).
8. Warren, P. *Nature* **308**, 492–493 (1984).
9. Bond, A. & Sparks, R. S. J. *J. geol. Soc. Lond.* **132**, 1–17 (1976).
10. Pichler, H. & Friedrich, W. L. in *Thera and the Aegean World II* (ed. Doumas, C.) 15–30 (Aegean World, London, 1980).
11. Heiken, G. & McCoy, F. Jr *J. geophys. Res.* **89**, 8441–8462 (1984).
12. Wilson, L. in *Thera and the Aegean World II* (ed. Doumas, C.) 31–35 (Aegean World, London, 1980).
13. Sparks, R. S. J., Brazier, S., Huang, T. C. & Muerdter, D. *Mar. Geol.* **54**, 131–167 (1983).
14. Simkin, T., Siebert, L., McClelland, L., Bridge, D., Newhall, C. & Latter, J. M. *Volcanoes of the World* (Hutchinson Ross, Stroudsburg, 1981).
15. Betancourt, P. P. *Archaeometry* **29**, 45–59 (1987).
16. Michael, H. N. in *Thera and the Aegean World I* (ed. Doumas, C.) 791–796 (Aegean World, London, 1978).
17. Michael, H. N. in *Thera and the Aegean World II* (ed. Doumas, C.) 245–246 (Aegean World, London, 1980).
18. Friedrich, W. L., Friberg, R. & Tauber, H. in *Thera and the Aegean World II* (ed. Doumas, C.) 241–243 (Aegean World, London, 1980).
19. Meulengracht, A., McGovern, P. & Lawn, B. *Radiocarbon* **23**, 227–240 (1981).
20. Pearson, G. W. & Stuiver, M. *Radiocarbon* **28**, 839–862 (1986).
21. Bruns, M., Levin, I., Münnich, K. O. & Hubberten, H. H. *Radiocarbon* **22**, 532–535 (1980).
22. Hammer, C. U., Clausen, H. B. & Tauber, H. *Radiocarbon* **28**, 284–291 (1986).
23. Hammer, C. U. *J. Glaciol.* **25**, 359–372 (1980).
24. Hammer, C. U. *et al.* *Am. Geophys. Union (AGU) Monograph* **33**, 90–94 (1985).
25. Lamb, H. H. in *Climate: Present, Past and Future* Vol. I 412–417 (Methuen, London, 1972).
26. Hammer, C. U., Clausen, H. B. & Dansgaard, W. *Nature* **288**, 230–235 (1980).
27. Hammer, C. U. *Jökull* **34**, 51–65 (1984).
28. Miller, T. P. & Smith, R. L. *Geology* **15**, 434–438 (1987).
29. LaMarche, V. C. & Hirschboeck, K. K. *Nature* **307**, 121–126 (1984).

Evidence from a New Zealand snail for the maintenance of sex by parasitism

Curtis M. Lively

Department of Zoology, University of Canterbury, Christchurch 1, New Zealand

The widespread occurrence of sexual reproduction is an important problem in evolutionary theory, because of the cost of producing males in dioecious organisms and the allocation of limited resources in coxeses^{1–4}. Here I compare the predictions of the major ecological hypotheses on the maintenance of sex by examining male frequency in *Potamopyrgus antipodarum*, a dioecious freshwater snail with both sexual and parthenogenetic populations⁵. The results do not support the view that sex is maintained by a variable physical environment^{6–9}, but they are consistent with the idea that sex is favoured by selection resulting from host–parasite interactions^{10–14}.

Two of the three major ecological hypotheses on the maintenance of sex emphasize variation in the physical environment ('sex' is used here to mean biparental reproduction). The temporal variation hypothesis suggests that sex is favoured by unpredictable environmental changes that occur between generations. This is because sexual populations^{6,7}, or families², would seem more likely to have at least some members capable of survival and reproduction following such changes. The related spatial variation hypothesis^{8,9} claims that sex is favoured by intraspecific competition in an environment which is heterogeneous in space, especially if there is a history of competition among sibs^{2,15}. This idea is related to models of genetic polymorphism, which show that the maintenance of multiple alleles

is favoured by resource competition in a spatially variable environment¹⁶⁻¹⁸.

The third hypothesis, called the Red Queen hypothesis, emphasizes the importance of frequency-dependent selection resulting from interspecific interactions, such as those between hosts and their parasites¹⁰⁻¹⁴. The advantage of sex (and recombination¹⁹⁻²⁰) under this model is, for the host, the potential for the production of rare offspring that may escape infection and, for the parasite, the capacity to track these genotypes as they become common.

The three hypotheses might apply to *P. antipodarum* as follows. The temporal variation hypothesis predicts that sex should be more common in streams than in lakes, assuming that streams are more unstable in time. This assumption is probably true in general²¹, and it has support in the particular case of New Zealand streams²².

The spatial variation hypothesis, in contrast, predicts that sex should be more common in lakes, making the assumptions that because of their greater stability, intraspecific competition is greater in lakes, and that lakes have multiple niches upon which different snail genotypes might specialize. This same prediction, however, might be arrived at in another way: because stream populations are typically more sparse and individuals are more likely to be isolated by floods, parthenogenesis should be more

Table 1 Trematode infection of snails from New Zealand streams and lakes

	Streams	Lakes	<i>H</i>	<i>P</i>
Proportion of males	2.39 (0.86)	12.75 (2.82)	9.708	0.002
<i>Microphallus</i> sp.*	2.22 (0.73)	7.85 (1.88)	12.654	<0.001
<i>Stegodexamene anguillae</i>	0.69 (0.22)	0.73 (0.19)	1.536	0.215
<i>Coitocaeum anaspidis</i>	0.80 (0.41)	0.72 (0.31)	1.568	0.211
<i>Telogaster opisthorchis</i>	0.71 (0.28)	0.72 (0.18)	3.097	0.079
Monostome cercariae (four spp.)	0.46 (0.19)	1.79 (0.82)	6.300	0.012
Xiphidiocercariae (two spp.)	1.81 (0.61)	0.47 (0.21)	1.587	0.208
Total parasites†	7.32 (1.08)	13.89 (2.39)	9.708	0.002

Means with standard error in parentheses for the untransformed percentages of males and individuals infected by the different trematode groups for *P. antipodarum* collected from New Zealand streams (*N* = 29) and lakes (*N* = 22). *H* is the Kruskal-Wallis statistic for a nonparametric, one-way ANOVA on ranked data; *P* is the probability of *H*.

* Undescribed metacercariae tentatively assigned to *Microphallus*; all other species reside in the snail as cercariae.

† Total includes all the listed species plus some rare and unidentified species.

common in streams (if the probability of isolation is high enough), as it ensures fitness for females in isolation^{4,23}. (Note that this idea, called the reproductive assurance hypothesis, is concerned with the loss of sex when there is a risk of isolation. It relies on the other hypotheses for the maintenance of sex when there is no such risk.)

Finally, the Red Queen hypothesis predicts that sex should be most common in populations where there has been a history of parasitism. *P. antipodarum* is the intermediate host for at least fourteen species of digenetic trematodes²⁴, so assuming that present-day infections are representative of past selection, the Red Queen hypothesis predicts that sex should be positively related to the frequency of trematode-infected snails. Note that the trematodes occur in the snails as larvae, which can not be directly transferred among snails. (Direct transfer from parents to their progeny enhances the expected potential for parasites to maintain sex in their hosts¹³.) Most of the larvae go through a second intermediate host (a crustacean), and then mature in a vertebrate. This indirect life cycle means that infection levels

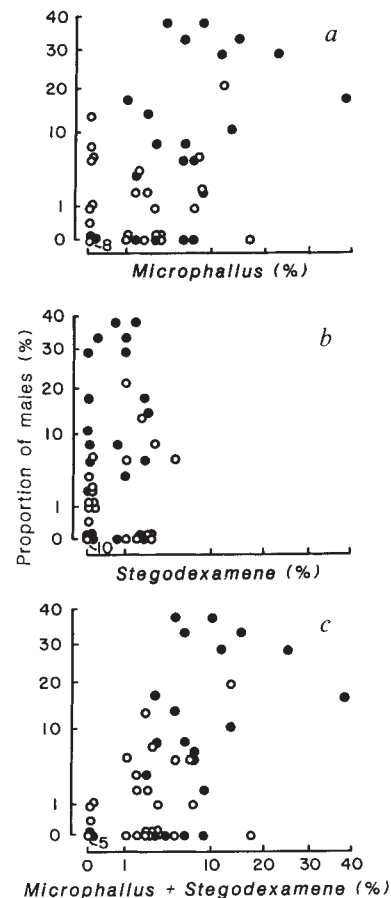


Fig. 1 The proportion of males in stream (○) and lake (●) populations of *Potamopyrgus antipodarum* plotted against the percentages of individuals infected with *Microphallus* sp. (a), *Stegodexamene anguillae* (b), and either *Microphallus* or *Stegodexamene* (c). The axes were back transformed from arcsine $\sqrt{x}$. The lake samples are from lakes Rotorua and Tawarewa on New Zealand's North Island and from lakes Forsythe, Ellesmere, Sara, Grasmere, Hawdon, Ianthe, Paringa, Kaniere, Mapourika, Wombat, Wahapo, Brunner, Haupiri, Lady, Poerua, Alexandrina, McGregor, Murray, Ida and Evelyn on the South Island. All stream samples are from streams (mostly unnamed) on the South Island.

are not expected to be a simple function of snail density²⁵, so that, for example, a dense population could be parasite-free if the final host is absent.

To test these hypotheses, *P. antipodarum* were collected from lake margins and in streams throughout New Zealand, particularly on the South Island, and a minimum of forty individuals (usually 100) were sexed and examined for the presence of trematodes under a dissecting microscope (infected individuals rarely contained more than one trematode species). In cases where I sampled the same site more than once, and where Winterbourn²⁴ sampled two of the same lakes previously, I used the average among samples as the single datum for that site. The frequency of males was used as an estimator of sexual reproduction in each population. (A recent isozyme analysis of *P. antipodarum* confirms that males make a genetic contribution to the offspring of sexual females, and that reproduction in parthenogenetic females is ameiotic²⁶.)

A significantly higher mean frequency of males was observed in lakes than in streams (Table 1), in direct opposition to the prediction of the temporal variation hypothesis (see also refs 9 and 27). In addition, the average proportion of snails infected by the most common trematode species (tentatively assigned to *Microphallus*) was also greater in lakes. These findings are concordant with the spatial variation and reproductive assurance hypotheses, which both predicted more males in lakes, and the

Table 2 Results of stepwise multiple regression analysis

Step	Source	Regression MS (d.f.)	Residual MS (d.f.)	F	Total R ²
No variables forced					
1	<i>Microphallus</i>	0.589 (1)	0.032 (49)	18.56*	0.275
2	<i>Stegodexamene</i>	0.589 (2)	0.028 (48)	14.98*	0.384
Habitat forced					
1	Habitat	0.510 (1)	0.033 (49)	15.28*	0.238
2	<i>Microphallus</i>	0.369 (2)	0.029 (48)	12.58*	0.344
3	<i>Stegodexamene</i>	0.301 (3)	0.026 (47)	11.38*	0.421

Summary table for the stepwise multiple regression of per cent males of *P. antipodarum* against habitat and the trematode groups listed in Table 1. MS, mean square. In the first run, none of the variables were forced³⁷, and they were entered by the program (SPSS) in order of their unadjusted partial R^2 values until none of the remaining values was significant. In the second run, habitat was forced as the first variable entered, and the remaining values were entered as just described. All continuous variables were transformed (arcsine $\sqrt{x}$) before the analysis.

* $P < 0.001$; all other steps not significant ($P > 0.10$).

Red Queen hypothesis, which predicted a positive relationship between male frequency and infection level.

These hypotheses were compared using a stepwise multiple regression of per cent males (dependent variable) against habitat and the percentages of snails infected by the trematode groups listed in Table 1 (independent variables). The percentage of snails infected with metacercariae of *Microphallus* sp. and cercariae of *Stegodexamene anguillae* entered first and second, respectively; none of the remaining variables were significant. In a second analysis, I forced habitat as the first variable entered, and a similar result was obtained: *Microphallus* entered next, followed by *Stegodexamene* (Table 2). Hence male frequency is significantly correlated with the frequency of infected individuals, even when the effects of habitat are first removed; but the converse is not true. This indicates that parasites are more important than habitat, which supports the Red Queen hypothesis (see also refs 28–30).

The Red Queen hypothesis gains additional support from the relationships between male frequency and infection levels within habitats (Table 3). In lakes, male frequency is significantly correlated with the frequency of snails infected by *Microphallus* (Fig. 1a); and in streams, male frequency is significantly correlated with the frequency of snails infected by *Stegodexamene* (Fig. 1b). Interestingly, when the frequencies of individuals infected by these two trematodes are added together (Fig. 1c), the correlations improve (Table 3).

These results suggest a role for trematodes, especially *Microphallus*, in the maintenance of sex in *P. antipodarum*, but they are not conclusive by themselves. We require further that the correlation between sex ratio and *Microphallus* is not an artefact. One way in which the correlation could be an artefact is if *Microphallus* differentially infects males, creating a situation where the greater the frequency of males, the greater the frequency of infected individuals. To test this idea, I compared the frequencies of males and females infected by *Microphallus* in Lake Alexandrina (chosen because it was my most extensively

sampled site). The proportion of males (13 out of 96) infected by *Microphallus* was not significantly different than that for females (83 out of 576): $G = 0.05$, $P > 0.80$. Hence the correlation is probably not an artefact of differential infectivity of the sexes (but it could result from an unknown third force that controls both sex ratio and parasite load).

We also require to know that reproductive success of the snail is negatively affected by trematode infection, and that *P. antipodarum* has a heritable resistance to such infection. It is clear that the trematodes could exert a strong selective force on the snail: following successful infections, the trematodes reproduce asexually and replace most of the viscera and gonad³¹ of the snail. This is especially true of *Microphallus*, which encysts in the snail host. It is not presently known whether the snail has a genetically based resistance to infection by the trematodes, but such resistance is known for snails serving as intermediate hosts for schistosomes^{32–34}. Also unknown are the proximate determinants of the female-biased sex ratios. Facultative parthenogenesis is one possibility, but it would seem unlikely to account for the sex ratio variation among lakes (unless parthenogenesis is conditional on cues other than the simple presence of males). Another possibility (supported by Wallace³⁵) is that sexual females, producing a 1:1 sex ratio, coexist with obligately parthenogenetic females³ (see also ref. 36). Experiments to test whether parthenogenesis in *P. antipodarum* is facultative or obligate are in progress.

In conclusion, the Red Queen hypothesis is the best supported of the hypotheses considered here, because of the positive correlations between the sex ratio of the snail and infection by trematodes both within and between habitats. This result identifies parasitism, in nature, as a potential source of frequency-dependent selection favoring the maintenance of sex.

I thank D. Blair, L. F. Delph, S. Deblock, A. Graesser, D. G. Lloyd, J. McKenzie, M. J. McKone, J. Maynard Smith, B. Peckarsky, N. Phillips, R. Rowe and M. Winterbourn. This work was supported by a post-doctoral fellowship from the University Grants Committee of New Zealand.

Received 9 April, accepted 1 June 1987.

- Maynard Smith, J. in *Group Selection* (ed. Williams, G. C.) 163–175 (Aldine Atherton, Chicago, 1971).
- Williams, G. C. *Sex and Evolution* (Princeton University Press, 1975).
- Maynard Smith, J. *The Evolution of Sex* (Cambridge University Press, 1978).
- Lloyd, D. G. in *Evolutionary Biology* Vol. 13 (eds Hecht, M. K., Steere, W. C. & Wallace, B.) 69–111 (Plenum, New York, 1980).
- Winterbourn, M. *Malacologia* 10, 283–321 (1970).
- Fisher, R. A. *The Genetical Theory of Evolution* (Oxford University Press, 1930).
- Muller, H. J. *Am. Nat.* 66, 118–138 (1932).
- Ghiselin, M. T. *The Economy of Nature and the Evolution of Sex* (University of California Press, Berkeley, 1974).
- Bell, G. *The Masterpiece of Nature* (University of California Press, Berkeley, 1982).
- Jaenike, J. *Evol. Theory* 3, 191–194 (1978).
- Hamilton, W. D. *Oikos* 35, 282–290 (1980).
- Levin, D. *Am. Nat.* 109, 437–451 (1975).
- Rice, W. R. *Am. Nat.* 121, 1317–1320 (1983).
- Bremermann, H. J. *J. theor. Biol.* 87, 641–702 (1980).
- Price, M. V. & Wasser, N. M. *Evolution* 36, 35–43 (1982).
- Levine, H. *Am. Nat.* 87, 331–333 (1953).
- Maynard Smith, J. 1970. *Am. Nat.* 140, 487–490 (1970).
- Strobeck, C. *Am. Nat.* 108, 152–156 (1974).
- Hamilton, W. D. in *Population Biology of Infectious Diseases* (eds Anderson, R. M. & May, R. M.) 269–296 (Springer, New York, 1982).
- Maynard Smith, J. in *The Evolution of Sex: An Examination of Current Ideas* (eds Levin, B. R. & Michod, R. E.) (Sinuaer Sunderland, Massachusetts, in the press).
- Hynes, H. B. N. *The Ecology of Running Waters* (Liverpool University Press, 1970).
- Winterbourn, M. J., Rounick, J. S. & Cowie, B. N. *Z. J. mar. Freshwat. Res.* 15, 321–328 (1981).
- Lloyd, D. G. in *The Evolution of Sex: An Examination of Current Ideas* (eds Levin, B. R. & Michod, R. E.) (Sinuaer Sunderland, Massachusetts, in the press).
- Winterbourn, M. J. *Mauri Ora* 2, 17–30 (1973).
- May, R. M. & Anderson, R. M. *Nature* 280, 455–461 (1979).
- Phillips, N. thesis Univ. Auckland (1987).
- Jaenike, J. & Selander, R. K. *Am. Zool.* 19, 729–737 (1979).
- Ellstrand, N. C. & Antonovics, J. *Evolution* 39, 657–666 (1985).
- Antonovics, J. & Ellstrand, N. C. *Evolution* 38, 103–115 (1984).
- Schmitt, J. & Antonovics, J. *Evolution* 40, 818–829 (1986).
- McArthur, C. P. & Featherston, D. W. *N. Z. J. Zool.* 3, 35–38 (1976).
- Richards, C. S. *Parasitology* 70, 231–241 (1975).
- Newton, W. L. *Expl Parasit.* 2, 252–257 (1953).
- Richards, C. S. & Merritt, J. W. *Am. J. trop. Med. Hyg.* 21, 425–434 (1972).
- Wallace, C. J. *Malac. Soc. Aust.* 4, 263 (1980).
- Case, T. J. & Taper, M. L. *Evolution* 40, 366–387 (1986).
- Sokal, R. R. & Rohlf, F. J. *Biometry* 2nd edn (Freeman, New York, 1981).

Table 3 Product-moment correlation coefficients

	<i>Microphallus</i>	<i>Stegodexamene</i>	<i>Microphallus</i> + <i>Stegodexamene</i>
Streams	0.179	0.361*	0.375*
Lakes	0.492*	0.126	0.515*
Streams and lakes	0.524**	0.251	0.583**

Product-moment correlation coefficients for the frequencies of males and the frequencies of *P. antipodarum* infected by *Microphallus* and *Stegodexamene* in New Zealand lakes ($N = 22$) and streams ($N = 29$).

* $P < 0.05$; ** $P < 0.001$.

Signal clipping by the rod output synapse

David Attwell*, Salvador Borges†, Samuel M. Wu‡ & Martin Wilson†

* Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK

† Department of Zoology, University of California, Davis, California 95616, USA

‡ Cullen Eye Institute, Baylor College of Medicine, Houston, Texas 77030, USA

The properties of synapses between retinal neurons make an essential contribution to early visual processing. Light produces a graded hyperpolarization in photoreceptors, up to 25 mV in amplitude¹, and it is conventionally assumed that all of this response range is available for coding visual information. We report here, however, that the rod output synapse rectifies strongly, so that only potential changes within 5 mV of the rod dark potential are transmitted effectively to postsynaptic horizontal cells. This finding is consistent with the voltage-dependence of the calcium current presumed to control neurotransmitter release from rods². It suggests functional roles for the strong electrical coupling of adjacent rods³ and the weak electrical coupling of adjacent rods and cones⁴. The existence of photoreceptor coupling resolves the apparent paradox that rods have a 25 mV response range, while signals greater than 5 mV in amplitude are clipped during synaptic transmission. We predict that the strengths of rod-rod and rod-cone coupling are quantitatively linked to the relationship between the rod response range and the synapse operating range.

The input-output relation for rod to horizontal cell transmission was studied in the tiger salamander retina using three independent approaches (Figs 1–3), all of which demonstrated strong synaptic rectification. An approximation to the steady-state input-output relation was obtained by recording simultaneously the voltage response of a rod (Fig. 1*a*) and a horizontal cell (Fig. 1*b*) to bright light flashes. Such stimuli evoke a rod hyperpolarization which recovers very slowly (over about a minute) to the dark potential, and a cone hyperpolarization which outlasts only briefly the period of application of the light⁴. Consequently, although horizontal cells in this retina receive synaptic input from both rods and cones^{5,6}, the slow return of the horizontal cell potential to its dark level after a flash reflects the recovery of the rod potential, and rod to horizontal cell transmission can be studied by plotting the horizontal cell potential as a function of rod potential during this period (Fig. 1*c*). The input-output relation shows strong rectification: for rod hyperpolarizations less than 2 mV in amplitude the gain is high (7.2 ± 4.9 , mean $\pm$ s.d., in six rod-horizontal cell pairs in the flat-mounted retina), but hyperpolarizations of 5 mV or greater almost saturate the synapse. The whole input-output curve was approximately exponential, with a gain that decreased *e*-fold for each 2.1 mV of rod hyperpolarization (mean value 2.10 ± 0.44 mV in six cell pairs).

To investigate the transmission of rapid rod polarizations, we recorded the response of a horizontal cell to the injection of current into a rod (Fig. 2). Current injection into one rod evokes voltage changes in surrounding rods, due to electrical coupling of the rods. These responses are more transient at greater distances from the site of current injection because of signal shaping by voltage-gated currents in the rod membrane³. In all the rods, the voltage change at the onset of a hyperpolarizing current pulse is larger than that at the start of a depolarizing pulse, because of outward rectification in the rod current-voltage (*I*-*V*) relation³. Nevertheless, the response in postsynaptic horizontal cells was larger at the onset of depolarizing current into a rod, indicating a lower synaptic gain for transmitting rod hyperpolarizations.

The largest postsynaptic signal in Fig. 2 is evoked at the

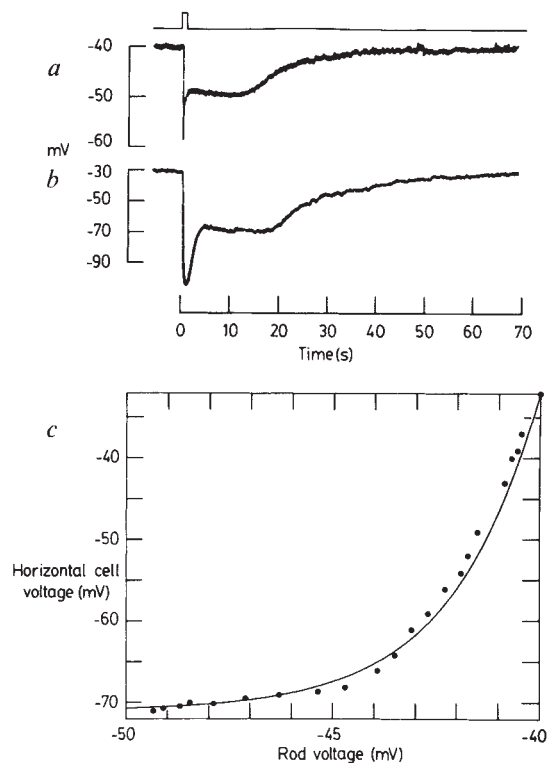


Fig. 1 Steady-state input-output relation for rod to horizontal cell transmission. Responses of a rod (*a*) and a horizontal cell (*b*) to a 2-mm spot of 500-nm light (3×10^4 photons $\mu\text{m}^{-2} \text{s}^{-1}$, timing shown by top trace). Light suppresses an inward current flowing into the rod outer segment¹² and thus evokes a rapid (<100 ms) hyperpolarization of nearly 20 mV, followed by a rapid (<300 ms) partial repolarization of about 10 mV which is produced by voltage-gated currents in the rod inner segment^{2,3}. After this, and in the absence of further illumination, the rod stays hyperpolarized for a period depending on the light intensity (up to 30 s), before slowly repolarizing to the dark potential as the light-sensitive inward current in the rod outer segment recovers. The voltage change occurring in the horizontal cell is produced by synaptic input from both rods and cones^{5,6}. The bright lights used here to evoke prolonged rod voltage tails evoke a hyperpolarization in cones during the application of light, but the cones rapidly repolarize on removal of the illumination¹³ (apart from a small hyperpolarization remaining, due to current flow into cones through gap junctions with rods⁴). Consequently, cone input produces an initial component of horizontal cell hyperpolarization (during the first 3 s of *b*) superimposed on the hyperpolarization produced by rod input. However, the slow recovery of the horizontal cell and rod voltages after $t = 10$ s in these records reflects only the recovery of the light-sensitive current in rods. *c*, Plot of the horizontal cell voltage, V_H , as a function of the rod voltage, V_R , for $t > 10$ sec in *a* and *b*. Smooth curve has the form

$$V_H - V_{H,\text{dark}} = A(\exp\{(V_R - V_{R,\text{dark}})/k\} - 1)$$

where: the horizontal cell and rod dark potentials are $V_{H,\text{dark}} = -32$ mV and $V_{R,\text{dark}} = -40$ mV; $A = 39$ mV; and k is 2.1 mV. This expression must slightly underestimate the nonlinearity of synaptic transmission from rods, as some of the rod signal reaching horizontal cells leaves the receptor layer via the cone output synapse⁷.

Methods. Simultaneous recordings were made from a rod and a horizontal cell in the isolated flat-mounted retina, or in slices of retina¹⁴. Data in this figure are from a flat-mounted retina. Microelectrodes were filled with 0.5- or 3-M K-acetate (resistance 300–900 M Ω). Illumination was applied along the long axis of the photoreceptors, either through the objective lens (for flat-mounted retinas) or from the side of the recording chamber (for slices). An infrared television system was used to view the preparation⁴, to avoid bleaching the photoreceptor pigment. Preparations were perfused with a solution containing (mM): NaCl 104; KCl 2; MgCl₂ 1; CaCl₂ 2; glucose 5; HEPES 5; pH 7.6.

Fig. 2 Investigation of rod to horizontal cell transmission by current injection into rods. Top, the response of rods in the flat-mounted retina to the injection of ± 1 and ± 2 nA, for approximately 2 s, into the rod shown impaled with an electrode in the inset (from ref. 3: the voltage response at the site of current injection was recorded with a second electrode in the same rod). Similar results are obtained for experiments on retinal slices. Signals spread to rods around the injected rod via gap junctions coupling adjacent rods. Centre to centre spacing of adjacent rods is about $16 \mu\text{m}$. The rod response becomes more transient as it spreads through the rod network because of shaping by voltage-gated currents in the rod membrane³. Note different voltage scales for different rods. Bottom, response of a horizontal cell in a retinal slice to injection of ± 1 and ± 2 nA into a rod. Note timescale is different from that for rod data. The responses are largely transient, perhaps indicating that they are dominated by synaptic input from the large number of rods undergoing a transient voltage change (see top records), rather than by input from the rod where current is injected. The horizontal cell hyperpolarization produced by injection of current into a rod is much smaller than the peak light-induced response because current injection polarizes only a small fraction of the rods in the horizontal cell receptive field. Similar results were obtained for 10 horizontal cells in the flat-mounted retina, except that the amplitude of the responses was approximately half that found in retinal slices (26 cells). This difference probably reflects the higher input resistance of horizontal cells in retinal slices, where half the horizontal cell dendritic tree is removed by the slicing process.

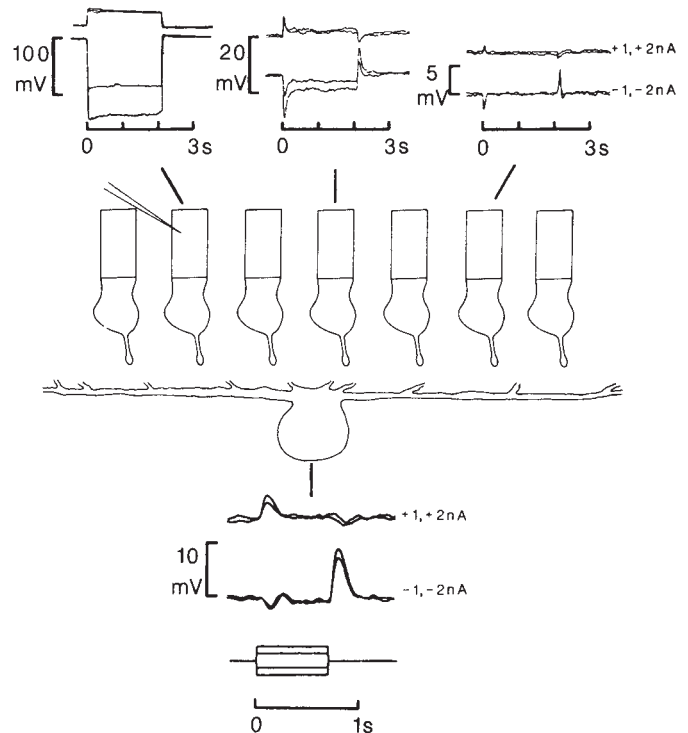
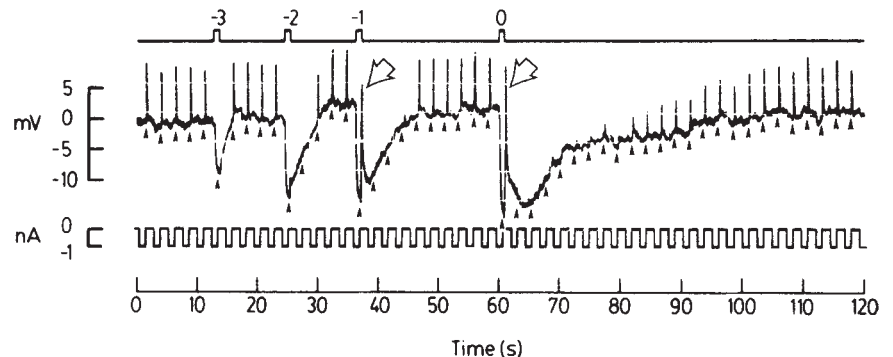


Fig. 3 The effect of rod hyperpolarization by light, on signal transmission evoked by injecting current into rods. A hyperpolarizing current pulse (-1 nA, bottom trace) was repeatedly injected into a rod in a retinal slice, evoking a postsynaptic horizontal cell response like that in Fig. 2. At the end of each -1 nA pulse (marked by a solid arrow head), a transient depolarization is evoked in the horizontal cell potential (middle trace: 0 mV denotes the dark potential of -40 mV). White light flashes (top trace) were then applied to hyperpolarize the photoreceptors and hence the horizontal cells. Flash intensity was incremented by a factor of 10 for each subsequent flash (-3 , -2 , -1 , 0 denote the $\log_{10}$ (intensity), with 0 equivalent to approximately $8,000 \text{ photons } \mu\text{m}^{-2} \text{ s}^{-1}$ at 520 nm wavelength incident on the rods). As the flash is made brighter, the rod and horizontal cell voltage tails occurring after the flash are prolonged, and a transient depolarization (open arrows) produced by cone input at the end of each flash^{6,15} becomes larger (this transient depolarization is larger than in Fig. 1 because white, rather than green light was used as a stimulus). The response to current injection into the rod is suppressed by the light-induced hyperpolarization of the rod network, although the rod network response to current injection is only slightly affected by illumination¹⁶. This suppression cannot be due to a decrease in horizontal cell membrane resistance associated with hyperpolarization, because the horizontal cell resistance is increased by illumination⁶. Furthermore, hyperpolarizing the horizontal cell with current through the recording electrode increased the amplitude of the response to current injected into rod (data not shown), probably due to displacement of the horizontal cell potential further from the reversal potential for the synaptic input⁶.



termination of hyperpolarizing pulses into the rod. A large transient depolarization occurs, that probably reflects the transient depolarization past the resting potential occurring in rods other than that where current is injected (Fig. 2, top). Figure 3 shows that this transient depolarization occurring in the horizontal cell is suppressed by hyperpolarization of the rod network with broad-field illumination. During the slow recovery of the horizontal cell potential after illumination, the current-induced postsynaptic signal reappears only when the horizontal cell potential (and hence the rod potential) has returned to within a few millivolts of its dark level. Thus, for transient changes in rod voltage, as for the slower presynaptic signals investigated in Fig. 1, we conclude that the gain of rod to horizontal cell transmission is greatly reduced by rod hyperpolarization.

Rectification in synaptic transmission from rods is consistent with the properties of the calcium current in these cells, which is activated at potentials above -45 mV, and increases e -fold

with each 6 mV of depolarization². Thus, if neurotransmitter release is controlled by the calcium concentration in the rod synaptic terminal, hyperpolarization from the dark potential (-40 mV) to -45 mV will decrease the calcium influx and reduce transmitter release, but further hyperpolarization should produce no further change in transmitter output. Rod signals larger than 5 mV in amplitude will thus be clipped during synaptic transmission. (On this basis, a similar rectification is expected for transmission from rods to bipolar cells: for the experiment of Fig. 2, this is indeed found⁷.) Cooperativity in the action of the rod transmitter on its postsynaptic receptors⁸ may also contribute to the synaptic rectification we observe.

In the salamander retina adjacent rods are strongly electrically coupled to each other by gap junctions^{3,9}. Electrical coupling is generally assumed to improve the signal to noise ratio in the rod network: by coupling rods together the voltage noise generated by independent current fluctuations in different rods will be reduced, but the signal produced by broad-field illumination

will be unaffected¹⁰. However, for the situation where noise is most important—the detection of a single photon—the electrical coupling of rods makes the signal to noise ratio worse¹¹. Our demonstration of rectification in the rod output synapse suggests an alternative reason for the existence of rod-rod coupling. During localized illumination (for example of only one rod), electrical coupling will spread the presynaptic signal over a number of rods, each of which undergoes a small hyperpolarization and thus has its synapse operating at a high gain. This will produce a larger postsynaptic response in a cell receiving input from all the coupled rods than if there was no coupling.

Thus, for localized illumination rod-rod coupling partly negates the effect of signal clipping produced by the rod output synapse. For broad-field illumination, however, signal spread between rods is absent because all the rods undergo the same voltage change, and large signals will be clipped by the rod output synapse. This suggests a function for the weak electrical coupling known to exist between rods and cones in the salamander retina^{4,9}. Rod hyperpolarizations greater than the level which are clipped by the rod synapse are transmitted to adjacent cones (reduced in amplitude by a factor of about five⁷), and will modulate transmitter release from the cone output synapse. In this way the dynamic range of the rod system for broad-field stimuli will be extended beyond the limit imposed by the rod synapse properties.

The existence of rod-rod and rod-cone coupling suggests a resolution of the apparent paradox that, whereas rods have a voltage response range extending over 25 mV, only the first 5 mV or so of that range can usefully code visual information for transmission through the rod synapse. Consider the effect of rod-rod coupling first. If the 5 mV operating range of the output synapse is to be used fully for coding different intensities of light falling on just one rod, then bright light falling on only one rod must generate at least a 5 mV response in that cell. This will only occur if the peak response to broad-field illumination, for which no current flows from one rod to another, is arranged to have an amplitude much larger than the 5 mV operating range of the synapse: larger, in fact, by the ratio of input resistances of an isolated rod and a coupled rod in the retina, that is, about 7 in the tiger salamander retina³. Conversely, if the maximum response to broad-field illumination were set at only 5 mV (in order to match the range of rod responses produced by broad-field stimuli to the synapse coding range) then for localized illumination, which would generate responses up to 5/7 mV in amplitude, the coding capability of the synapse would be greatly underutilized. Second, consider the influence of rod-cone coupling in extending the operating range of the rod system for broad-field stimuli. Since rod signals are reduced by a factor of about 5 in transmission to cones⁷, a broad-field rod response of about 25 mV is needed in order to use the whole coding range of the cone output synapse (assuming this to rectify like the rod synapse).

Thus, these considerations lead to the surprising conclusion that the output synapse properties and the strengths of rod-rod and rod-cone coupling are inherently related to each other. We predict, as is found experimentally, that the ratio of the peak rod response amplitude to the synapse operating range (25 mV/5 mV) should have approximately the same value as the ratio of input resistances of an isolated rod and a rod in the retina (700 M Ω /100 M Ω), and the same value as the factor by which broad-field rod responses are attenuated during transmission to cones (about 5). In a particular animal species, the value of these three ratios must depend on a combination of biophysical constraints, for example, the voltage range of activation of calcium channels which imposes a general limitation on synaptic transmission, with constraints related to the nature of the visual information being processed, such as requirements for spatial resolution determining the degree of rod-rod coupling.

We thank Marc Tessier-Lavigne for useful discussion of these

results. This work was supported by the MRC, the Wellcome Trust, NIH, NSF, NATO and the Central Research Fund of London University.

Note added in proof: J. H. Belgum & D. R. Copenhagen (*J. Physiol., Lond.* (in the press)) have compared rod, horizontal and bipolar cell voltages in the toad retina and found an *e*-fold decrease in rod synaptic gain with each 2 mV of rod hyperpolarization.

Received 7 April; accepted 11 June 1987.

1. Bader, C. R., MacLeish, P. R. & Schwartz, E. A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 3507–3511 (1978).
2. Bader, C. R., Bertrand, D. & Schwartz, E. A. *J. Physiol., Lond.* **331**, 253–284 (1982).
3. Attwell, D. & Wilson, M. J. *J. Physiol., Lond.* **309**, 287–315 (1980).
4. Attwell, D., Wilson, M. & Wu, S. M. *J. Physiol., Lond.* **352**, 703–737 (1984).
5. Hanani, M. & Vallergha, S. J. *J. Physiol., Lond.* **298**, 397–405 (1980).
6. Marshall, L. M. & Werblin, F. S. J. *J. Physiol., Lond.* **279**, 321–346 (1978).
7. Attwell, D. Q. *J. exp. Physiol.* **71**, 497–536 (1986).
8. Shiells, R. A., Falk, G. & Naghsineh, S. *Proc. R. Soc. Lond. B* **227**, 121–135 (1986).
9. Custer, N. V. *J. comp. Neurol.* **151**, 35–56 (1973).
10. Falk, G. & Fatt, P. *Handbook of Sensory Physiology* Vol. VII/1 (ed. Dartnall, H. J. A.) 235–239 (Springer, Heidelberg, 1972).
11. Lamb, T. D. & Simon, E. J. *J. Physiol., Lond.* **263**, 257–286 (1971).
12. Baylor, D. A., Lamb, T. D. & Yau, K. W. *J. Physiol., Lond.* **288**, 589–611 (1979).
13. Attwell, D., Werblin, F. S. & Wilson, M. J. *J. Physiol., Lond.* **328**, 259–283 (1982).
14. Werblin, F. S. *J. Physiol., Lond.* **280**, 449–470 (1978).
15. Normann, R. A. & Perlman, I. J. *J. Physiol., Lond.* **286**, 509–524 (1979).
16. Attwell, D., Wilson, M. & Wu, S. M. *Brain Res.* **343**, 79–88 (1985).

Deletion of genes on chromosome 1 in endocrine neoplasia

C. G. P. Mathew, B. A. Smith, K. Thorpe, Z. Wong*, N. J. Royle*, A. J. Jeffreys* & B. A. J. Ponder

Institute of Cancer Research, Clifton Avenue, Sutton, Surrey SM2 5PX, UK

* Department of Genetics, University of Leicester, University Road, Leicester LE1 7RH, UK

Recent studies have identified normal cellular DNA sequences which are lost in the development of embryonal^{1–6} and adult^{7,8} tumours. These tumours are thought to arise after a primary mutation in one allele of such a sequence is followed by loss of its normal homologue. In familial cases, the primary mutation is transmitted in the germ line. The secondary mutation may involve a substantial loss of chromosomal material and thus lead to identification of the site of the inherited mutation. We have examined constitutional and tumour genotypes of medullary thyroid carcinomas and pheochromocytomas which develop in the dominantly inherited cancer syndrome multiple endocrine neoplasia type 2 (MEN2)⁹ to locate the predisposing gene in this syndrome. We observed deletion of a hypervariable region of DNA on the short arm of chromosome 1 in seven out of fourteen tumours. Analysis of the parental origin of the deleted allele in two families showed that it was derived from the affected parent in one case, which suggests that the deletion does not reflect the site of the inherited mutation in MEN2. The deleted region is distal to the breakpoint commonly detected in neuroblastomas, which share with the tumours of MEN2 embryological origin from neuroectoderm.

We chose to search for loss of heterozygosity by using a panel of locus-specific minisatellite probes^{10,11} isolated from a human genomic library using the minisatellite probes 33.15 and 33.6 (ref. 12). These locus-specific probes each detect multiple alleles with a high degree of heterozygosity, using a limited repertoire of restriction enzymes^{10,11}. Analysis of constitutional and tumour genotypes with the probe λ MS1 (*DIS7*), which detects a hypervariable minisatellite region on chromosome 1, showed that 7 of 14 informative pairs had a loss or marked reduction in intensity of one of their two alleles in the tumour DNA (Fig. 1).

Table 1 Loss of heterozygosity on chromosome 1 in endocrine neoplasias

Probe	λ MS1	λ MS32	λ MS8	λ MS29	p λ g3	pMS51	λ MS43
Map symbol	<i>D1S7</i>	<i>D1S8</i>	<i>D5S43</i>	—	<i>D7S22</i>	—	<i>D12S11</i>
Enzyme	<i>Hinf</i> I	<i>Mbo</i> I	<i>Hinf</i> I	<i>Hinf</i> I	<i>Hinf</i> I	<i>Hae</i> III	<i>Hinf</i> I
Chromosome	1p33-35	1q*	5	6	7	11	12
Patient							
Phaeochromocytoma							
RB (S)	12	—	U	U	12	12	12
JC (2A)	1	12	12	U	12	12	12
DH (S)	2	12	12	12	12	12	—
LP (S)	1	12	12	12	12	12	12
JS (NF)	2	12	12	12	12	—	12
DK (S)	12	—	—	—	—	—	—
Medullary thyroid carcinoma							
JJ (S)	12	12	12	U	12	12	12
MM (2A)	12	12	12	12	12	—	12
SP (2A)	2	12	12	12	U	12	12
JP (2A)	U	12	12	12	12	—	12
MS (2A)	12	—	12	12	12	—	12
DW (2A)	2	12	U	12	12	U	12
RW (S)	12	—	12	U	12	12	12
EM (2B)	2	—	U	12	—	—	12
CB (2A)	12	—	—	—	—	—	12

Human Gene Map symbols for λ MS29 and pMS51 have not yet been assigned. MEN type in parentheses after patient's initials: S, sporadic; 2A, MEN type 2A; NF, von Recklinghausen's neurofibromatosis; 2B, MEN type 2B (see Fig. 1 legend for details). The tumour DNA genotype is shown for all cases where the lymphocyte DNA showed heterozygosity: '12' indicates heterozygosity, although multiple alleles are observed with these probes; '1' or '2' indicates tumour DNA which retained only the allele of higher or lower relative molecular mass respectively. 'U' indicates that the lymphocyte DNA was homozygous and therefore uninformative and '—' that it was not tested.

Methods. Tumour and leukocyte DNA was digested with restriction enzymes, fractionated by agarose gel electrophoresis, transferred to nylon membranes and hybridized to the probes as in Fig. 1. All the probes used were isolated using the minisatellite core probes λ 33.15 and 33.6 (ref. 12); details of their characterization have been described elsewhere (refs 10, 11 and A.J.J. *et al.*, unpublished data). The regional assignment of the probe marked * is provisional. Probe λ MS1 has been regionally localized by *in situ* hybridization (N.J.R. *et al.*, unpublished data). The copy number of the λ MS1 allele that was retained in the tumours was determined by hybridizing the same filters with the probe p λ g3 on chromosome 7, and comparing the ratio of the intensity of this λ MS1 allele and a p λ g3 allele in the tumour DNA with that obtained for the same pair in leukocyte DNA.

These included four of six phaeochromocytomas and three of eight medullary thyroid carcinomas (MTCs) from both familial and non-familial cases. The chromosomal specificity of this loss was tested with probes from five other chromosomes (Table 1), and in no case was loss of heterozygosity observed. The data in Table 1 also indicate that the loss of one λ MS1 allele could not be explained by degradation of high relative molecular mass tumour DNA because in two cases it was the high molecular mass allele ('1') that was retained, and the tumours did not show the loss of allele 1 when tested with other probes. The extent of loss of hybridization signal from the affected λ MS1 allele was analysed by densitometry of autoradiographs in which film response had been linearized by pre-flashing¹³. The signal was reduced by 73% (in phaeochromocytoma patient DH, here abbreviated as P/DH), 66% (P/LP), 84% (P/JC), 69% (P/JS), 70% (in MTC patient SP, or MTC/SP), 88% (MTC/EM) and 53% (MTC/DW) respectively. The remaining hybridization signal from the affected allele may result from contaminating non-tumour cells present in the surgical specimen (see Fig. 1 legend). Alternatively, the tumours may be a mosaic of cells with a normal karyotype and cells with a loss or alteration of chromosome 1. The tumour specificity of the allele loss was analysed by typing eight bladder cancer leukocyte/tumour pairs with the λ MS1 probe. All eight pairs were informative, and none showed loss of an allele in the tumour DNA (data not shown).

Several different mechanisms could account for the loss of the homologous locus in the tumour DNA¹⁴. We have attempted to distinguish among these mechanisms firstly by typing the leukocyte/tumour pairs with λ MS32, which is also located on chromosome 1, but not linked to λ MS1 ($z > -2$ at $\theta > 0.35$, where z is the lod score for a recombination fraction θ). None of the nine informative pairs showed loss of heterozygosity with this probe (Table 1) including six patients who had lost heterozy-

gosity with λ MS1. Thus only part of chromosome 1 was lost or altered in the tumours, and this probably involves the short arm as the provisional regional localization of λ MS1 and λ MS32 is 1p33-35 and 1q respectively (ref. 11 and unpublished data). We have also compared the hybridization signals of the retained λ MS1 alleles with those obtained when the same filters were hybridized with a probe from a different chromosome (see Table 1 for details). The copy number of this allele relative to the corresponding allele in leukocyte DNA was 1.15 (P/JS), 0.90 (MTC/SP), 1.1 (P/LP) and 1.06 (MTC/DW). These data indicate that the retained allele is present in the hemizygous state, and are consistent with deletion of part of one copy of chromosome 1, rather than mitotic recombination, loss and reduplication, or gene conversion.

The question arises whether the deletions observed on chromosome 1 reflect the locus of the inherited mutation in the MEN2 syndrome. If this were so, the allele inherited from the unaffected parent in familial cases would be deleted in the tumour, as in heritable cases of retinoblastoma¹⁵. We were able to test this in two families with MEN2 (Fig. 2). In the family of MTC/SP, the λ MS1 allele from the affected father (lane c) is retained in the tumour (lane b), as predicted. However, in P/JC the allele retained in the tumour (lane d) was inherited from the unaffected father (lane b). This unexpected result suggests that the deletion on chromosome 1 may not reflect the site of the primary inherited mutation in MEN2. Preliminary analysis of the segregation of λ MS1 alleles in a large MEN2 kindred indicates a lack of linkage between these two loci ($z > -2$ at $\theta > 0.11$, our unpublished data). The significance of this observation is twofold. Firstly it implies that studies of leukocyte/tumour genotypes in other inherited cancer syndromes should include analysis of the parental origins of the retained and deleted alleles. It also suggests that the second event in these tumours may not necessarily be the loss of the

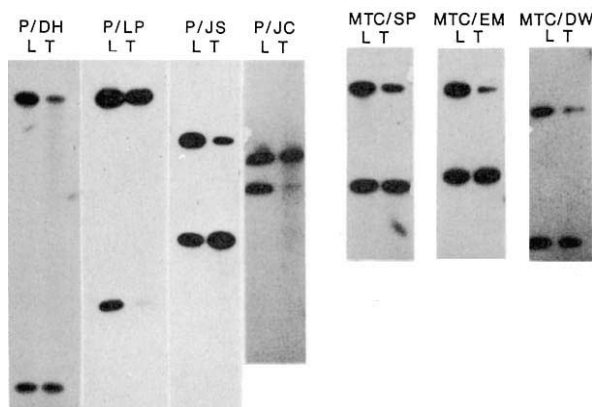


Fig. 1 Loss of heterozygosity of a locus on human chromosome 1 in endocrine neoplasias. DNA from leukocytes (L) and tumours (T) of patients with pheochromocytomas (P) or medullary thyroid carcinomas (MTC) was typed with the locus-specific minisatellite probe λ MS1 (ref. 11), which detects a hypervariable DNA segment on chromosome 1. See text for notation designating patients.

Methods. Cases studied included patients with MEN2A (MTC, pheochromocytoma and parathyroid hyperplasia), MEN2B (a more aggressive form than type 2A, also characterized by mucosal neuromas and a Marfanoid habitus), Von Recklinghausen's neurofibromatosis and sporadic cases with no family history of these tumours. Surgical tumour specimens were snap-frozen in liquid N_2 immediately after removal and the diagnosis confirmed by histological analysis of haematoxylin-eosin stained frozen sections before isolation of DNA. (Some tumours showed a small amount of contamination with normal connective tissue of low cellularity.) Tumour tissue was ground with a pestle and mortar, and suspended in 10 vols TNE buffer (10 mM Tris-HCl, pH 8, 0.1 M NaCl, 1 mM EDTA). The suspension was incubated at 37 °C for 3 h in 1% SDS and 0.1 mg ml⁻¹ proteinase K. Tumour DNA was extracted with phenol and chloroform, and spooled onto a glass rod after addition of 2 vol ethanol. DNA was extracted from peripheral blood as previously described²¹. DNA samples were digested with *Hin*II (Amersham), fractionated by gel electrophoresis and transferred to nylon filters. The probe insert fragment was separated from λ vector fragments by electrophoresis and oligolabelled directly in agarose to a specific activity of 1×10^9 c.p.m. μ g⁻¹ (ref. 22). Southern hybridization, post-hybridization washing and autoradiography were performed as described previously^{11,23}.

normal homologue of the gene in which the primary mutation occurred. For example, the chromosome 1 deletion could reflect the loss of a tumour suppressor gene which results in the expression of a dominantly acting MEN2 gene. This might require mutation or loss of one or both copies of the putative suppressor gene. Evidence for the existence of a suppressor gene on chromosome 1 has been provided by studies of hybrids between normal human fibroblasts and a human fibrosarcoma cell line¹⁶ or transformed baby hamster kidney cells¹⁷, in which re-expression of transformation was correlated with loss of human chromosome 1. Alternatively, a mutation in one of the several oncogenes located on chromosome 1 (ref. 18) followed by loss of its wild-type allele, could lead to cooperation between the mutant oncogene and the MEN2 gene in tumorigenesis.

Chromosome 1 changes have been reported in many tumours¹⁹, but these most commonly result in duplication of long-arm material. Over 70% of neuroblastomas, however, showed loss of material on the short arm of chromosome 1, with the most frequent breakpoint being 1p32 (ref. 20). Interestingly, the locus detected by λ MS1 (1p33-35) is just distal to this breakpoint, and MTCs, pheochromocytomas and neuroblastomas are all derived from neural crest cells during embryogenesis. It is therefore possible that deletions of genes on the

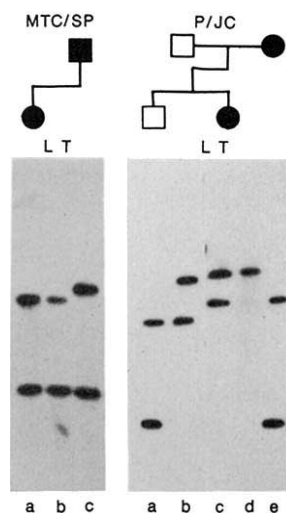


Fig. 2 Inheritance of the retained and deleted λ MS1 alleles in two families with MEN2A. Left, patient MTC/SP: the allele inherited from the patient's affected father has been retained in the tumour. Right, patient P/JC: the allele inherited from the patient's unaffected father has been retained in the tumour. The familial relationship of these individuals was confirmed by rehybridization of the filters with a minisatellite probe from chromosome 7 (data not shown). The correct identification of samples from the parents of JC was confirmed by the observation that tumour DNA from the affected mother (MTC/DW) had the same alleles as her leukocyte DNA with all seven minisatellite probes. The λ MS1 alleles were analysed as described in Fig. 1. Symbols: filled squares, affected male; filled circle, affected female; open square, unaffected male.

short arm of chromosome 1 may reflect a common mechanism in the development of these neoplasias.

We thank Dr I. Summerhayes for DNA samples from the bladder tumour/leukocyte pairs; Dr J. Sloane for the histological analyses and the following clinicians who provided tissue for this study: J. Brown, A. Chilvers, L. Gracey, J. Hale, D. Heath, M. Maisey, J. Monson, A. Nash, A. O'Meara, P. Rosswick, P. Semple, A. Young. This research is supported by a programme grant to the Institute of Cancer Research by the Cancer Research Campaign and the MRC and by a grant to A.J.J. from the MRC. A.J.J. is a Lister Institute Research Fellow.

Note added in proof: The MEN2 gene has recently been mapped to chromosome 10 (see refs 24 and 25).

Received 3 June; accepted 17 June 1987.

1. Cavenee, W. K. *et al.* *Nature* **305**, 779-784 (1983).
2. Koufos, A. *et al.* *Nature* **309**, 170-172 (1984).
3. Orkin, S. H., Goldman, D. S. & Sallan, S. E. *Nature* **309**, 172-174 (1984).
4. Reeve, A. E. *et al.* *Nature* **309**, 174-176 (1984).
5. Fearon, E. R., Vogelstein, B. & Feinberg, A. P. *Nature* **309**, 176-178 (1984).
6. Koufos, A. *et al.* *Nature* **316**, 330-334 (1985).
7. Fearon, E. R. *et al.* *Nature* **318**, 377-380 (1985).
8. Seizinger, B. R., Martuza, R. L. & Gusella, J. F. *Nature* **322**, 644-647 (1986).
9. DeLellis, R. A. *et al.* *Int. Rev. exp. Path.* **28**, 163-215 (1986).
10. Wong, Z., Wilson, V., Jeffreys, A. J. & Thein, S. L. *Nucleic Acids Res.* **14**, 4605-4615 (1986).
11. Wong, Z., Wilson, V., Patel, I., Povey, S. & Jeffreys, A. J. *Ann. hum. Genet.* (in the press).
12. Jeffreys, A. J., Wilson, V. & Thein, S. L. *Nature* **314**, 67-73 (1985).
13. Laskey, R. A. & Mills, A. D. *Eur. J. Biochem.* **56**, 335-341 (1975).
14. Solomon, E. *Nature* **309**, 111-112 (1984).
15. Cavenee, W. *et al.* *Science* **228**, 501-503 (1985).
16. Benedict, W. F. *et al.* *Cancer Res.* **44**, 3471-3479 (1984).
17. Stoler, A. & Bouck, N. *Proc. natn. Acad. Sci. U.S.A.* **82**, 570-574 (1985).
18. Povey, S., Morton, N. E. & Sherman, S. L. *Cytogenet. Cell Genet.* **40**, 67-106 (1985).
19. Atkin, N. B. *Cancer Genet. Cytogenet.* **21**, 279-285 (1986).
20. Gilbert, F. *et al.* *Cancer Res.* **44**, 5444-5449 (1984).
21. Vandenplas, S. *et al.* *J. med. Genet.* **21**, 164-172 (1984).
22. Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **137**, 266-267 (1984).
23. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
24. Mathew, C. G. P. *et al.* *Nature* **328**, 527-528 (1987).
25. Simpson, N. E. *et al.* *Nature* **328**, 528-530 (1987).

A linked genetic marker for multiple endocrine neoplasia type 2A on chromosome 10

C. G. P. Mathew*, K. S. Chin*, D. F. Easton*, K. Thorpe*, C. Carter*, G. I. Liou†, S.-L. Fong†, C. D. B. Bridges‡, H. Haak‡, A. C. Nieuwenhuijzen Kruseman‡, S. Schifter§, H. H. Hansen§, H. Telenius||, M. Telenius-Berg¶ & B. A. J. Ponder*

* Institute of Cancer Research, Haddow Laboratories, Clifton Avenue, Sutton, Surrey SM2 5PX, UK

† Department of Ophthalmology, Baylor College of Medicine, Houston, Texas 77030, USA

‡ Department of Endocrinology, University Hospital of Leiden, Leiden, The Netherlands

§ Department of Nuclear Medicine, Aarhus Municipal Hospital, 8000 Aarhus C, Denmark

|| Hassleholm Hospital, Hassleholm, Sweden

¶ Department of Internal Medicine, Faculty of Medicine, Safat, Kuwait

Multiple endocrine neoplasia type 2A (MEN2A) is an autosomal dominantly inherited cancer syndrome characterized by medullary carcinoma of the thyroid, pheochromocytoma and hyperparathyroidism. Almost all gene carriers can be detected by screening tests before the age of 40 (ref. 1), but the nature and location of the predisposing gene are unknown. Simpson *et al.*² recently reported preliminary evidence for linkage between the DNA probe p9-12A on chromosome 10 and MEN2A. We now report linkage between the MEN2A locus and the interstitial retinol-binding protein gene, which is located on chromosome 10p11.2-q11.2 (ref. 3).

The use of DNA probes to detect linkage between a restriction fragment length polymorphism (RFLP) and a mutated locus has led to the successful mapping of inherited disorders such as Duchenne muscular dystrophy⁴, Huntington's chorea⁵, polycystic kidney disease⁶ and cystic fibrosis^{7,8}. In view of the need for a closely linked genetic marker for MEN2A, several linkage studies have been undertaken using both classical genetic polymorphisms⁹ and RFLPs¹⁰⁻¹⁴, and a recent report indicates that about 32% of the genome has been excluded from linkage to this disorder². We have tested 23 probes for linkage to MEN2A but found no evidence for close linkage (ref. 15; C.G.P.M. *et al.*, manuscript in preparation). A preliminary report² did, however, give a positive lod score of 1.7 (weak evidence of linkage) at a recombination fraction of 0.15, with an anonymous probe on chromosome 10. To determine which regions on the genome would be most likely to contain the MEN2A gene, all available linkage results were used to calculate a probability density map for the gene by Bayesian analysis¹⁶, assuming no genetic heterogeneity and prior probabilities for each chromosome proportional to its physical length. This analysis gave a posterior probability of 0.75 for the MEN2A gene being located on chromosome 10. In view of this we have analysed chromosome 10 probes for linkage to MEN2A.

We used a cDNA probe from the interstitial retinol-binding protein (IRBP) gene³ on chromosome 10 which detects an RFLP with the restriction endonuclease *Bgl*II (ref. 17). DNA from potentially informative members of nine MEN2A families were typed for the *Bgl*II RFLP. Only one branch of a large Swedish kindred and one Danish kindred were informative for the marker, which has a minor allele frequency of 0.15 (ref. 17). The segregation of the *Bgl*II alleles is shown in Fig. 1. The informative pedigrees and typing of the *Bgl* II alleles are shown in Fig. 2. The MEN2A gene is inherited with the less common

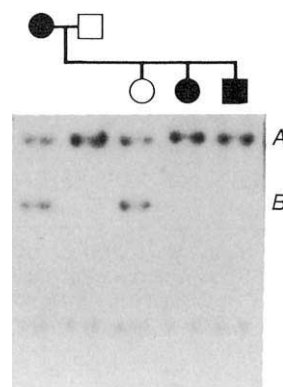


Fig. 1 Segregation of the *Bgl*II RFLP with the MEN2A locus in part of a family with two affected children (solid symbols) and one unaffected child (open symbol). Squares, males; circles, females. The RFLP alleles are A, of 6.3 kilobases (kb) and B, of 4.3 kb. There is a constant fragment of 2.5 kb.

Methods. DNA (5 µg) was digested with *Bgl* II, separated by electrophoresis on 0.6% agarose gels and transferred to Hybond N membranes (Amersham) by standard methods²⁰. The 2.2-kb complementary DNA inset (H.4IRBP) encoding human IRBP (ref. 3) was labelled to a specific activity of 1×10^9 d.p.m. µg⁻¹ by the random oligonucleotide priming reaction²¹ and hybridized as described²². Blots were washed to a final concentration of $0.1 \times$ SSC, 0.01% SDS at 65 °C ($1 \times$ SSC is 0.15 M NaCl, 0.015 M sodium acetate, pH 7).

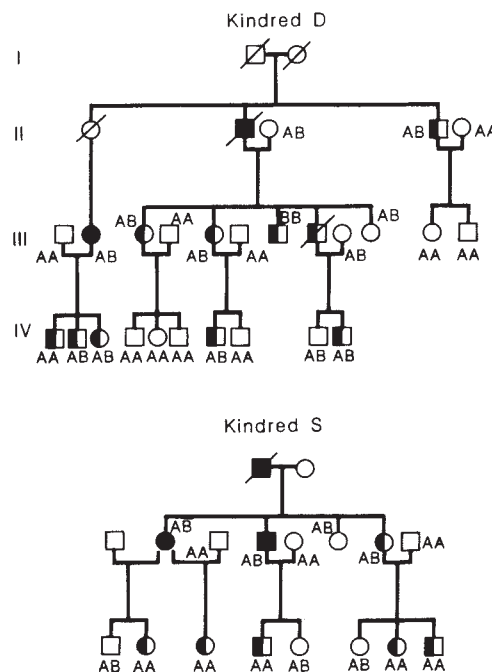


Fig. 2 Typing of the *Bgl*II RFLP in a Danish (D) and Swedish (S) kindred with typical multiple endocrine neoplasia type 2A, including medullary thyroid carcinoma (left part of symbol solid) and pheochromocytoma (right part of symbol solid). Squares, males; circles, females. All individuals at risk were screened by measurement of serum calcitonin after pentagastrin stimulation²³, and all unaffected individuals typed were over 25 years of age at the time of their last screen. The diagnosis of MEN2A was confirmed by histological examination of tissue obtained at surgery in all individuals scored as affected. There is clinical and histological evidence of parathyroid involvement in several members of other branches of the Swedish kindred, although none in the branch scored here. To date, no member of the Danish kindred has shown evidence of parathyroid involvement clinically or at thyroidectomy, and all have serum calcium values within the normal range.

Table 1 Lod scores at various recombination fractions (θ) for linkage between the loci MEN2A and IRBP

	θ						
Family	0	0.01	0.05	0.1	0.2	0.3	0.4
Kindred S	2.31	2.27	2.12	1.92	1.50	1.04	0.53
Kindred D	-2.31	1.32	1.75	1.71	1.30	0.75	0.26
Total	0.00	3.59	3.87	3.63	2.80	1.79	0.79

Likelihood computations were carried out using the LINKAGE program¹⁹, assuming 95% penetrance for the MEN2A gene¹, a gene frequency of 2×10^{-5} (ref. 15) and equal male and female recombination fractions.

B allele in kindred D, but with the common A allele in kindred S. The only recombinant detectable in the two kindreds is an affected male in generation IV of kindred D. The linkage relationship between the marker and the MEN2A locus is shown in Table 1. The maximum lod score between the locus defined by IRBP and the MEN2A locus is 3.88 at a recombination fraction of 0.04 (95% confidence interval 0.02 to 0.18). A lod score of 3.0 is accepted as strong evidence of linkage between two loci¹⁸.

The linkage data reported here establish that in at least some families the predisposing mutation in MEN2A maps relatively close to the centromere of chromosome 10. Because both informative families in this study originated from Scandinavia, the possibility of genetic heterogeneity cannot be excluded. We are screening the IRBP locus for more informative RFLPs to extend our analysis to families from different populations. Such markers would also enable us to establish the genetic distance between the IRBP and MEN2A loci more precisely. These studies should lead to the isolation of closely linked markers for the MEN2A gene, which could then be used for screening and genetic counselling in affected families. They will also provide the impetus for the identification and cloning of the predisposing gene itself.

We thank J. Edwards for advice and assistance with a computer-generated MEN2 exclusion map, and N. Simpson and K. Kidd for data in advance of publication. This work was supported by a programme grant to the Institute of Cancer Research from the Cancer Research Campaign and the MRC; and by project grants to the Cancer Research Campaign Medullary Thyroid Group. M. Telenius-Berg is supported by grants from the Swedish Medical Research Council, and S. Schifter by grants from the Danish Cancer Society. We thank numerous clinical colleagues for providing access to MEN2 families under their care.

Assignment of multiple endocrine neoplasia type 2A to chromosome 10 by linkage

N. E. Simpson*, K. K. Kidd||, P. J. Goodfellow*, H. McDermid*, S. Myers†, J. R. Kidd||, C. E. Jackson¶, A. M. V. Duncan*, L. A. Farrer||, K. Brasch*, C. Castiglione||, M. Genel#, J. Gertner#, C. R. Greenberg**, J. F. Gusella††, J. J. A. Holden§ & B. N. White‡

Departments of Biology*, Paediatrics†, Pathology‡, Psychiatry§, Queen's University, Kingston, Ontario K7L 3N6, Canada

|| Department of Human Genetics, Yale University School of Medicine, New Haven, Connecticut 06510, USA

¶ Department of Medicine, Henry Ford Hospital, Detroit, Michigan 48202, USA

Department of Pediatrics, Yale University School of Medicine, New Haven, Connecticut 06510, USA

** Department of Genetics, Health Sciences Center, Winnipeg, Manitoba R3E 0Z3, Canada

†† Neurology Department and Genetics Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA

Multiple endocrine neoplasia type 2A (MEN2A) is one of several kinds of cancers that appear to be inherited in an autosomally dominant fashion. We have assigned the MEN2A locus to chromosome 10 by linkage with a new DNA marker (*D10S5*). The linkage led us to investigate other chromosome 10 markers and demonstrate linkage between the disease locus and the interstitial retinol-binding protein (*IRBP*) gene¹. The *D10S5* locus was sublocalized to 10q21.1 by hybridization *in situ*² and the *IRBP* gene to p11.2→q11.2 with a secondary site at q24→q25¹. The linkages were established using 292 members of five families, three different restriction fragment length polymorphisms (RFLPs) at *D10S5* and two RFLPs recognized by the *IRBP* probe. The recombination frequencies from pairwise linkage analysis between the disease and two marker loci *D10S5* and *IRBP* were 0.19 and 0.11, with maximum lod scores of 3.6 and 8.0 respectively. Ordering of the three loci by multipoint analysis placed the *IRBP* gene approximately midway between the disease and *D10S5* loci.

The MEN2A syndrome is characterized by one or more of three types of tumours: medullary thyroid cancer (MTC), pheochromocytomas (PHEOs) and parathyroid adenomas, of which MTC is both the most frequent and most aggressive. MTC and its presumed precursor, C-cell hyperplasia, are commonly detected

Received 22 May; accepted 5 June 1987.

- Gagel, R. F. *et al.* *J. Paediat.* **101**, 941-946 (1982).
- Simpson, N. E. & Kidd, K. K. *Henry Ford Hosp. med. J.* (in the press).
- Liou, G. I. *et al.* *Som. Cell molec. Genet.* (in the press).
- Davies, K. E. *et al.* *Nucleic Acids Res.* **11**, 2303-2312 (1983).
- Gusella, J. F. *et al.* *Nature* **306**, 234-238 (1983).
- Reeders, S. T. *et al.* *Nature* **317**, 542-544 (1985).
- White, R. *et al.* *Nature* **318**, 382-384 (1985).
- Wainwright, B. J. *et al.* *Nature* **318**, 384-385 (1985).
- Kruger, S. D. *et al.* *Hum. Hered.* **36**, 6-11 (1986).
- Kidd, K. K. *et al.* *Henry Ford Hosp. med. J.* **32**, 262-265 (1986).
- Kidd, K. K., Castiglione, C. M., Pakstis, A. J. & Sparkes, R. S. *Genet. Epidemiol.* **3**, 195-200 (1986).
- Kidd, K. K. *et al.* *Hum. Hered.* **36**, 243-249 (1986).
- Goodfellow, P. J. *et al.* *Am. J. hum. Genet.* **37**, 890-897 (1985).
- Pakstis, A. J. *et al.* *Henry Ford Hosp. med. J.* (in the press).
- Carter, C. *et al.* *Cytogenet. Cell Genet.* (in the press).
- Sarfarazi, M. *et al.* *J. med. Genet.* **24**, (in the press).
- Liou, G. I. *et al.* *Nucleic Acids Res.* **15**, 3196 (1987).
- Morton, N. *Am. J. hum. Genet.* **7**, 277-318 (1955).
- Lathrop, G. M., Lalouel, J. M., Julier, C. & Ott, J. *Am. J. hum. Genet.* **37**, 482-498 (1985).
- Vandenplas, S. *et al.* *J. med. Genet.* **21**, 164-172 (1984).
- Feinberg, A. P. & Vogelstein, B. *Analyt. Biochem.* **137**, 266-267 (1984).
- Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
- Telenius-Berg, M. *et al.* *Eur. J. clin. Invest.* **7**, 7-16 (1977).

Table 1 Haplotype frequencies for the restriction fragment length polymorphisms at *D10S5* and for the *IRBP* gene

Locus	Observed haplotypes	Frequency	No. of chromosomes
<i>D10S5</i>	A1B1	0.58	102
	A1B2	not seen	
	A2B1	0.01	
	A2B2	0.41	
	A1B1C1	0.74	46
	A2B1C2	0.02	
	A2B2C2	0.24	
<i>IRBP</i>	Other haplotypes not seen		64
	A1B1	0.67	
	A1B2	0.20	
	A2B1	not seen	
	A2B2	0.13	

Table 2 Lod scores for pairwise linkage analyses

		Recombination fraction ($\theta_m = \theta_f$)									
Loci	Family	N1*	N2†	0.00	0.001	0.01	0.05	0.10	0.20	0.30	0.40
	S	69	56	-4.71	-3.60	-1.45	0.43	1.13	1.47	1.25	0.73
<i>MEN2A</i>	C	46	42	1.76	1.76	1.72	1.56	1.37	0.97	0.58	0.22
versus	K	28	19	0.03	0.33	1.09	1.74	1.87	1.66	1.18	0.57
<i>D10S5</i>	N	91	63	-6.53	-4.28	-2.46	-1.15	-0.66	-0.28	-0.12	-0.04
	W	58	41	-4.59	-4.27	-2.93	-1.54	-0.87	-0.24	0.01	0.07
	Total	292	221	-14.06	-10.06	-4.03	2.04	2.84	3.58	2.90	1.55
<i>MEN2A</i>	S	69	56	0.59	0.59	0.58	0.54	0.49	0.39	0.27	0.14
versus	C	46	41	-0.01	-0.01	0	0	0.01	0.01	0	0
<i>IRBP</i>	K	28	18	1.47	1.47	1.51	1.55	1.48	1.19	0.82	0.41
	N	91	63	-0.98	1.78	3.06	4.05	4.14	3.46	2.32	0.98
	W	58	42	-6.83	-5.68	-1.97	0.94	1.85	2.09	1.59	0.73
	Total	292	220	-5.76	-1.85	3.18	7.08	7.96	7.14	5.30	2.26

* N1, number of samples in analyses.

† N2, number of samples typed.

Lod scores were calculated using the program LIPED¹⁶ modified by an age of onset correction^{17,7}. The frequency of the disease allele used was 0.0005.

ted by a diagnostic serum calcitonin rise after stimulation with calcium infusion or pentagastrin. Early detection permits curative or prophylactic thyroidectomy. Identification of the allele or mutation for this dominantly inherited condition would eliminate annual or semiannual screening which is necessary because of the variable age of onset.

Previous attempts to map the *MEN2A* locus (*MEN2A*) had been unsuccessful³. A deletion on chromosome 20 had been reported in patients with *MEN2A*^{4,5} but the markers *D20S5* and *D2096* that mapped to the site of this deletion were not linked to the *MEN2A* locus in three large kindreds^{6,7}. Linkage studies of other markers had previously excluded 32% of the genome². We report here our linkage studies between the *MEN2A* locus and the two chromosome 10 markers *D10S5* and *IRBP*.

The five *MEN2A* families designated K⁸, C^{9,10}, S¹¹, N^{12,13} and W¹⁴ families have been previously described. The S family has presented with MCTs and parathyroid involvement, and K and C families with MCTs and PHEOs and the W and N families with the three clinical signs.

The probe p9-12A/2:dIII2.5 at the *D10S5* locus, its localization and polymorphisms have been described by McDermid *et al.*². The allele and haplotype frequencies (Table 1) for the RFLPs in *TaqI*, *HincII* and *DraI* digested DNAs were determined from unrelated spouses married into the five families with *MEN2A* and, for *TaqI*, unrelated individuals in non-*MEN2A* families in New Haven. Because the haplotypes for the *TaqI* and *DraI* polymorphisms were identical, except for one member of the W family, only the AB (*TaqI*, *HincII*) haplotypes were used in the linkage analysis.

Linkage between *D10S5* and the disease locus was detected (Table 2) with a maximum lod score (*z*) of 3.58 at a recombination frequency (θ) of 0.19. The one lod unit confidence interval¹⁵ was 0.09–0.33. We then tested other chromosome 10 markers in an attempt to find a closer marker and to resolve the question of possible heterogeneity, as the two families (N and W) showed no evidence of linkage to *D10S5* (Table 2). Linkage between the *MEN2A* locus and *D10S1* and *D10S4* has not been demonstrated (data not shown). The *BglII* RFLP for *IRBP* showed clear evidence of linkage in one family (N) but was largely non-informative in the other families. We screened the H4 clone of *IRBP* for RFLPs using 20 restriction enzymes and detected a high frequency polymorphism in *MspI* digests (Table 1) (J.R.K. *et al.* manuscript in preparation). The haplotypes for *IRBP* were informative for all but the C family and indicated linkage in the W family (Table 2). The positive lod scores for the *MEN2A* and *IRBP* loci in the two families (N and W) eliminate the possibility of heterogeneity in this sample of five

families. The maximum lod score for the *MEN2A* versus *IRBP* loci was 7.98 with $\hat{\theta} = 0.11$ (Table 2) with a one lod confidence interval of 0.04–0.21. The distance between *D10S5* and *IRBP* was estimated, from these and additional non-*MEN2A* families, as 8.7 centiMorgans (cM), corresponding to $\hat{\theta} = 0.09$ with a lod score of 9.01 and a one lod confidence interval of 0.03–0.17. Multipoint analysis (Fig. 1) places the *MEN2A* 12 cM outside and on the *IRBP* side of the *IRBP*–*D10S5* region.

This study demonstrates that the *MEN2A* locus is linked to two markers on chromosome 10. The two loci, however, are about 12 and 21 cM from the disease locus, both on the same side. Thus, the markers are not useful for precise predictions of the disease genotypes in families with this hereditary cancer. The assignment of the disease locus to a portion of chromosome 10 and the multiple RFLPs detected by the two probes, however, will enable a more detailed map of the region to be developed and closer markers to be identified to increase the understanding of the nature of the *MEN2A* mutation.

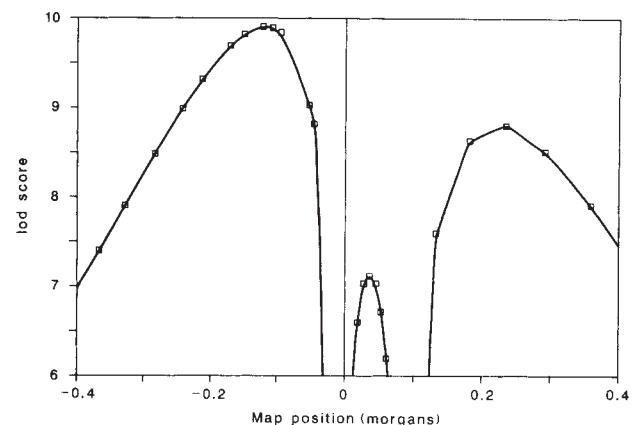


Fig. 1 Support for position of *MEN2A* with respect to *IRBP* and *D10S5*. The relative positions of the three loci were determined by multipoint analysis using the program LINKMAP from the package LINKAGE¹⁸. The age-dependent penetrance of *MEN2A* was defined as a step function for six age intervals⁷, corresponding to that used in the LIPED analyses. The position of *IRBP* was arbitrarily set at 0.00 and *D10S5* at 0.087 Morgans (M). The maximum lod score places the *MEN2A* locus at a map position of -0.122 M, with a one lod confidence interval of -0.250 M to -0.045 M. This position is favoured by odds of 575 to 1 and 13 to 1 over the other two peaks, at +0.037 M and an +0.232 M on the map, respectively. Our preliminary linkage studies of other chromosome 10 markers suggest that the order *IRBP*, *D10S5*, *MEN2A* is more unlikely than indicated in the figure.

We thank the families participating in this study, Drs. W. R. Ghent, Bruce Hamilton, A. J. Pakstis and Sergio Palmieri and Linda Anderson, Nancy Carson, Lois Cooke, Ivanka Franjkovic, Clive Morgan, Ellen Price and Cynthia Torony. We are grateful to Drs C. D. B. Bridges and G. I. Liou for the *IRBP* probe. This work was supported by the Canadian Medical Research Council (N.E.S.) and the Canadian National Cancer Institute (N.E.S. J.J.A.H. and B.N.W.), the US National Cancer Institute (K.K.K. and M.G., C.E.J.), the Hospital for Sick Children Foundation (A.M.V.D.), and by grants from the USNIH Clinical Research Center (M.G. and K.K.K.). A. M. V. Duncan was a Career Scientist of the Ontario Ministry of Health.

Received 18 May; revised 26 June; accepted 1 July 1987.

- Liou, G. I. *et al. Nucleic Acids Res.* **15**, 3196 (1987).
- McDermid, H. *et al. Nucleic Acids Res.* (in the press).
- Simpson, N. E. & Kidd, K. K. *Henry Ford Hosp. med. J.* (in the press).
- Babu, V. R., Van Dyke, D. L., & Jackson, C. E. *Proc. natn. Acad. Sci. U.S.A.* **81**, 2525-2528 (1984).
- Babu, V. R., Van Dyke, D. L., Flejter, W. C., Jackson, C. E. *Am. J. med. Genet.* (in the press).
- Goodfellow, P. J. *et al. Am. J. hum. Genet.* **37**, 890-897 (1985).
- Farrer, L. A. *et al. Cancer genet. Cytogenet.* (in the press).
- Partington, M. W., Ghent, W. R., Sears, E. V. P., & Simpson, N. E. *Can. med. Ass. J.* **124**, 403-410 (1981).
- Birr, A. R., Hogg, G. R., & Dubé, W. J. *Arch. Dermatol.* **113**, 1674-1677 (1977).
- Duncan, A. M. V. & Greenberg, C. R. *Cancer Genet. Cytogenet.* **22**, 109-112 (1986).
- Jackson, C. E. *et al. Ann. intern. Med.* **78**, 845-852 (1973).
- Hamilton, B. P., Landsberg L., & Levine, R. J. *Am. J. Med.* **65**, 1027-1032 (1978).
- Kruger, S. D. *et al. Hum. Hered.* **36**, 6-11 (1986).
- Keiser, H. R. Beaven, M. A., Doppman, J., Wells, Jr, S. & Buja, L. M. *Ann. Intern. Med.* **78**, 561-579 (1973).
- Ott, J. in *Analysis of Human Genetic Linkage* (Johns Hopkins University Press, Baltimore, 1985).
- Ott, J. *Am. J. hum. Genet.* **26**, 588-597 (1974).
- Morton, N. E. *Am. J. hum. Genet.* **7**, 277-318 (1955).
- Lathrop, G. M. *et al. Proc. natn. Acad. Sci. U.S.A.* **81**, 3443-3446 (1984).

Alteration of the quality of milk by expression of sheep β -lactoglobulin in transgenic mice

J. Paul Simons, M. McClenaghan & A. John Clark

Edinburgh Research Station, AFRC Institute of Animal Physiology and Genetics Research, Kings Buildings, West Mains Road, Edinburgh EH9 3JQ, UK

Milk contains a large amount of protein, most of which consists of a few major species synthesized in the mammary gland. The genes encoding these proteins are single-copy, and expressed during pregnancy and lactation¹. Although β -lactoglobulin (BLG) is the major protein in the whey of ruminants, it is not present in rodent milk. We have generated transgenic mice carrying the sheep BLG gene, and show that in such mice, BLG is specifically and abundantly expressed in the mammary gland during lactation. This results in a remarkable alteration of milk composition. These findings suggest that the manipulation of milk composition by gene transfer has considerable potential for the improvement of dairy animals.

The similarity of BLG to retinol binding protein has suggested that BLG is involved in vitamin A transport^{2,3}. It exists as a dimer of two identical subunits of 162 amino acids, each with relative molecular mass (M_r) of 18,000 (18K). In the mammary gland of lactating sheep, approximately 5% of the poly(A) RNA codes for BLG⁴. Four clones of the gene encoding sheep BLG were isolated from a genomic DNA library (A.J.C., unpublished data) constructed in the λ phage vector EMBL3⁵, using the BLG cDNA clone p931¹ as hybridization probe. Clone SS-1 contains 4 kilobases (kb) of DNA 5' to the cap site, the 4.9-kb BLG transcription unit and 7.3 kb of 3' flanking sequences (Fig. 1A), and was used to generate transgenic mice. The cloned BLG gene

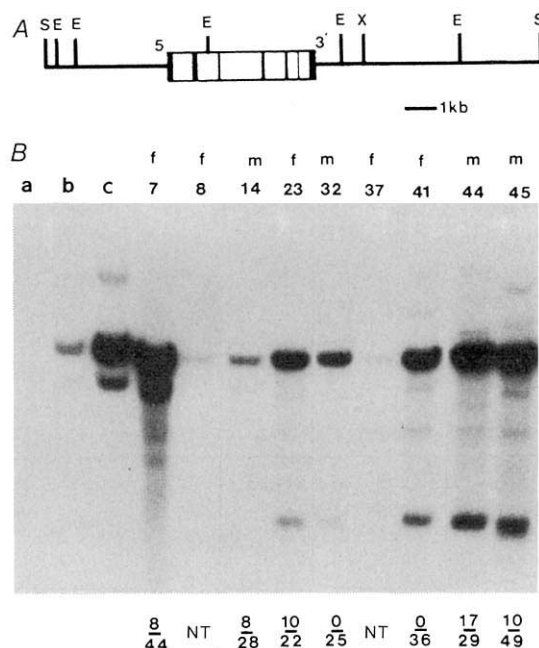


Fig. 1 Construction of transgenic mice carrying BLG gene segments. **A**, Structure of BLG insert of clone SS-1. Box, transcription unit of the gene; vertical lines within the box, positions and sizes of the seven exons. Restriction sites for *SalI* (S), *EcoRI* (E) and *XbaI* (X) are shown; both *SalI* sites are derived from the polylinker of λ EMBL 3. The DNA injected was the entire 16.2-kb *SalI* insert or the 10.5-kb *SalI*-*XbaI* fragment. **B**, Southern blot analysis of G0 mice. Numbered lanes contain samples from G0 mice. Lane a, control mouse DNA cleaved with *EcoRI*; lanes b, c, *EcoRI*-digested mouse DNA mixed with *EcoRI*-cleaved SS-1 in amounts equivalent to 1 and 10 copies per cell respectively. Mice 7, 8 and 14 are derived from embryos injected with the *SalI* fragment, the others from embryos injected with the *SalI*-*XbaI* fragment. The sex of each G0 mouse (m or f) is shown at top, and the frequency of transmission of the BLG sequences are shown below the lanes as a fraction of the progeny; NT, not tested for transmission.

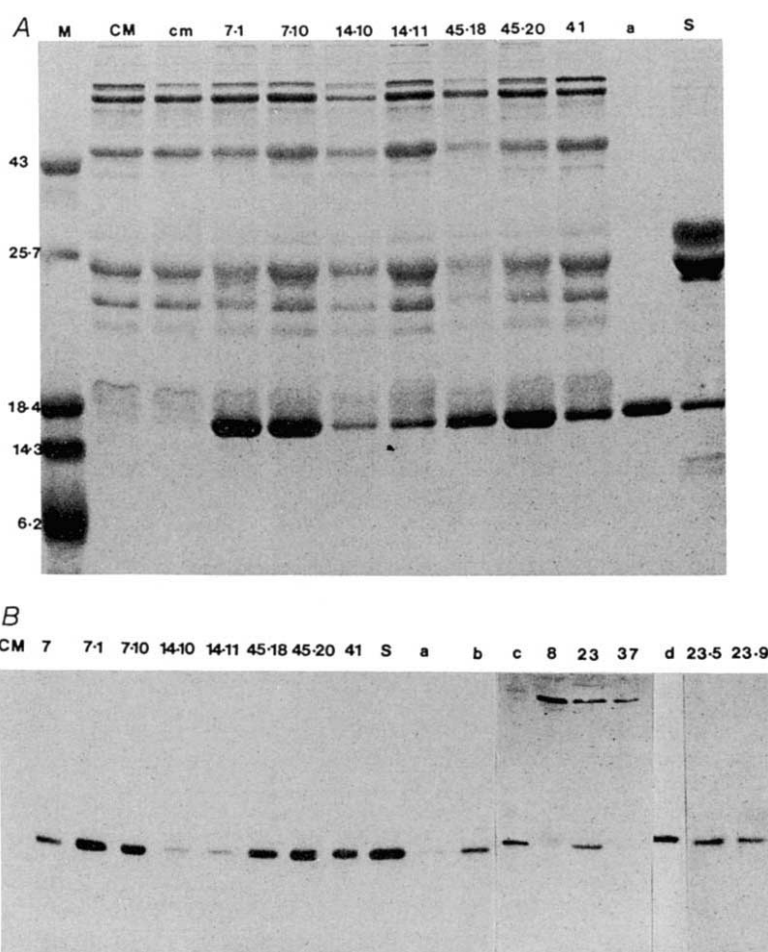
Methods. Before microinjection, SS-1 was cleaved with *SalI* or *SalI* and *XbaI* and 16.2- or 10.5-kb inserts comprising the gene and flanking sequences were isolated by preparative gel electrophoresis. Pronucleus stage eggs were obtained from superovulated C57BL/6 $\times$ CBA F₁ females, after mating with F₁ males. DNA (2 μ g ml⁻¹) was injected into either pronucleus using standard techniques; injected eggs were cultured overnight, cleaved embryos were transferred into the oviducts of pseudopregnant MF1 recipients. DNA was prepared from the tail of each mouse at weaning. *EcoRI*-cleaved DNA (10 μ g) was analysed by agarose gel electrophoresis, Southern blotting and hybridization with nick-translated pSS1tgX-S, a plasmid containing the 10.5-kb *SalI*-*XbaI* fragment of SS-1 subcloned into the vector pPoly1 (R. Lathe, personal communication).

was microinjected into fertilized eggs as the entire 16.2-kb insert (*SalI* fragment), or as a fragment with 5.7 kb removed from the 3' end (*SalI*-*XbaI* fragment). Of 325 eggs injected, 199 were transferred into pseudopregnant recipient females. Of the 46 'generation 0' (G0) offspring weaned, 16 were shown by Southern blot analysis to carry BLG sequences (6 with the *SalI* and 10 with *SalI*-*XbaI* fragment). The transgenic mice carried various numbers of copies of the BLG gene, ranging from <1 to ~20 copies per cell. Five G0 females and four G0 males (Fig. 1B) were selected for further analysis.

Seven G0 transgenic mice were tested for transmission (see Fig. 1B). Male mouse 32 and female mouse 41 failed to transmit the BLG sequences to their progeny. Of the five mice that did pass on the DNA, three (numbers 7, 14 and 45) did so at a frequency significantly below 50% ($P < 0.001$, 0.025 and 0.001 respectively), which may indicate mosaicism. One male mouse (number 44) was found to have incorporated the BLG gene on

Fig. 2 Sheep β -lactoglobulin in the milk of transgenic mice. **A**, Milk proteins stained with Coomassie Blue. Total milk proteins (equivalent of 0.2 μ l milk) from transgenic mice (numbered lanes), control C57BL/6 $\times$ CBA F₁ mouse (cm), control C57BL/6 mouse (CM) and sheep (S) were electrophoresed alongside BRL protein markers (M) and purified BLG (2.5 μ g, lane a). Size markers, M_r in thousands. **B**, Western blot of whey proteins. Whey proteins from transgenic mice (numbered lanes), control F₁ mouse (CM) and control sheep (S) were probed for sheep BLG. Solubilized whey (1 μ l is equivalent to 0.3 μ l milk) was loaded as follows: 0.1 μ l for mice 7.1, 7.10, 14.10, 14.11, 45.18, 45.20 and 41; 0.5 μ l for mouse 7; 1 μ l for the control mouse (CM) and sheep (S); 3 μ l for mice 23.5 and 23.9; 5 μ l for mice 8, 23 and 37. Standards: lanes a, b, c, and d were loaded with 100, 500, 2 and 100 ng of purified ovine BLG, respectively; note that a and b apply to the leftmost 12 lanes; c applies to the three lanes to the right of it, and d applies to the two lanes to the right of it. The high M_r band in tracks 8, 23 and 37 is due in each case to non-specific binding of the antiserum.

Methods. Milk was collected from nursing mice 11 days after parturition. The mother was separated from her pups and 3 h later injected intraperitoneally (i.p.) with 0.3 IU of oxytocin (Sigma); anaesthetic¹⁷ was administered after a further 10 min. Milk was collected by gentle massage of mammary glands and taken up in capillary tubes. Milk samples were diluted 1/5 in distilled water and defatted by centrifugation. To prepare whey, the caseins were precipitated by addition of 1 M HCl to pH 4.5 and removed by centrifugation. Total milk proteins were prepared for electrophoresis by solubilization of diluted, defatted milk in 4 vols loading buffer; whey proteins were precipitated in 5% trichloroacetic acid (TCA), washed in acetone and solubilized in loading buffer. Electrophoresis was according to Laemmli¹⁸ on 15% SDS polyacrylamide gels. **B**, Western blot analysis. After electrophoresis, proteins were transferred to nitrocellulose by the semi-dry blotting method¹⁹. Filters were blocked with horse serum and serially reacted with rabbit antibody against BLG and peroxidase labelled donkey antibody against rabbit immunoglobulin, and developed using diaminobenzidine as substrate.



the Y chromosome because all its sons and grandsons were transgenic, and all its daughters and grand-daughters were not.

Milk was obtained from lactating G0 transgenic female mice, and from transgenic progeny of each of the G0 mice which transmitted the BLG gene to daughters. Expression was initially assessed by SDS polyacrylamide gel electrophoresis (PAGE) and Coomassie Blue (CB) staining of total milk proteins. Sheep BLG was clearly evident in the milk of mice from line 7, line 14 and line 45, and of mouse 41, as a prominent band of relative molecular mass 18,000 (M_r 18K) comigrating with purified BLG and BLG in sheep milk (Fig. 2A). Milk from control mice gave a consistent pattern of bands; the only difference detected between milk proteins from control and transgenic mice was the presence of sheep BLG.

Western blotting of whey proteins confirmed the identity of the new protein as sheep BLG (Fig. 2B). This more sensitive analysis also allowed us to identify a further line of mice (line 23) which expresses BLG at a low level not detected by CB staining. Animals 8 and 37 did not detectably express BLG (less than 1 ng μ l⁻¹); both these animals were mosaic, having considerably less than one copy of the BLG gene per cell (Fig. 1B).

The concentrations of BLG in milk from animals of three abundantly expressing lines (7, 14 and 45), and from mouse 41 were estimated by densitometry of CB stained protein gels (Fig. 2A), using internal standards of purified BLG (Table 1). Levels as high as 23 mg ml⁻¹ (line 7) were found, more than five times the concentration estimated for BLG in sheep milk. Differences were found in BLG concentration between sisters of lines 14 and 45, reflecting the variation in the total protein content of individual milk samples (Fig. 2A). No differences were found between the duplicate mice of lines 7, 14 and 45 in

the staining intensity of the BLG bands relative to total staining in each sample (Table 1). This suggests that the qualitative composition of milk is constant in each line of transgenic mice. Further experiments are under way to quantitate precisely the protein content of milk from each of the lines and assess the effects of BLG expression on the rates of synthesis and concentration of endogenous milk proteins. The change in the composition of milk of transgenic mice was not accompanied by any apparent deleterious side effects, either to the lactating transgenic females or to the pups suckling their milk.

The tissue distribution of BLG expression was investigated in duplicate mice of lines 7, 14 and 45. RNA was prepared from a variety of tissues and analysed by Northern blotting (Fig. 3). Abundant BLG transcripts were detected only in mammary gland RNA, although after long exposures, extremely low levels of hybridization were detected in the other tissues analysed. Within the limits of resolution, mature BLG messenger RNA transcripts were found to be the same size as those in RNA isolated from sheep mammary gland (~800 nucleotides). Quantitative dot blotting (Table 1) showed that the mammary RNA of mice from lines 7 and 45 contained three to four times as much BLG RNA as mammary RNA from a lactating ewe.

The hormonal regulation of mammary development and milk secretion is similar in mice and sheep, although not identical^{6,7}. The major difference is that prolactin is required for the maintenance of lactation in mice, but not in sheep. The *cis*-acting sequences determining mammary expression of the BLG gene seem to be correctly interpreted in mice, despite both the absence of an equivalent gene, and the species differences in regulation. It is not yet known which sequences mediate mammary-specific expression, although both the DNA segments used to generate

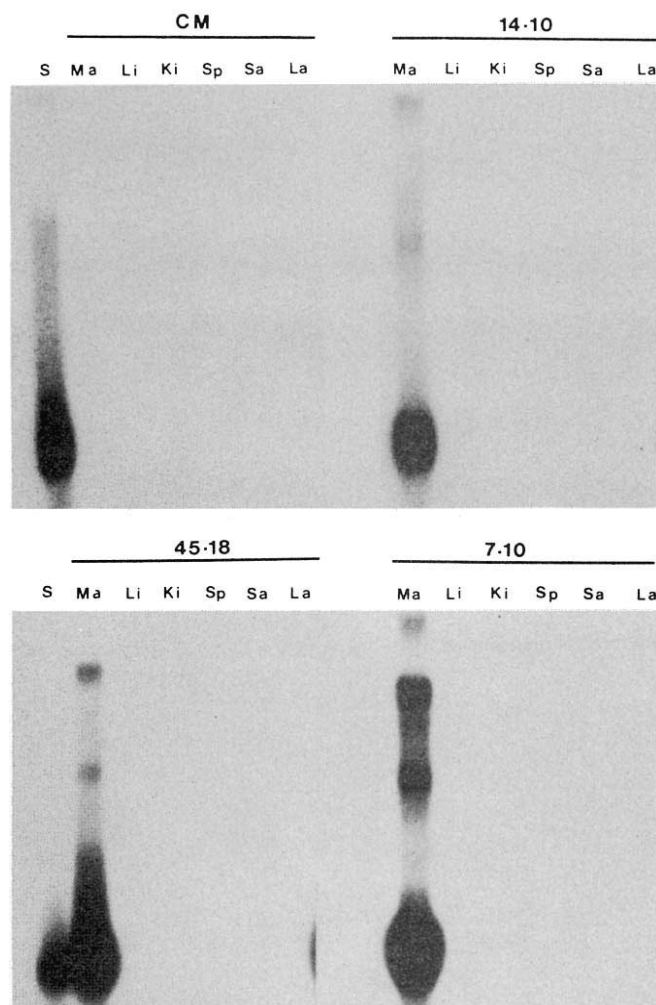


Fig. 3 The sheep BLG gene is expressed specifically in the mammary glands of transgenic mice. Northern blot analysis of total RNA from transgenic mice (numbers 7.10, 14.10 and 45.18) and control C57BL/6 mouse (CM). The tissues analysed were mammary glands (Ma), kidney (Ki), spleen (Sp), liver (Li), salivary glands (Sa) and lacrimal glands (La). Lanes S, RNA from the mammary gland of a lactating ewe. Duplicate transgenic mice (7.1, 14.11 and 45.20) gave essentially identical results.

Methods. RNA was prepared from lactating mice 11 days after parturition²⁰. Each lane contains 5 µg of total RNA, fractionated on denaturing 1% formaldehyde agarose gels. The RNA was transferred to Hybond N membranes (Amersham) and probed with nick-translated p931, as described²¹.

transgenic mice gave efficient tissue specific expression, showing that no essential regulatory sequences were present in the 5.7-kb region 3' to the *Xba*I site.

Transgenic mice carrying the sheep BLG gene produce milk with a significantly altered protein composition. It has been suggested that gene transfer might be used for the direct genetic improvement of domestic livestock⁸⁻¹⁰. Transgenic sheep and pigs have been produced carrying metallothionein-growth hormone (MT-GH) fusion genes in the hope of increasing growth rate and/or feed efficiency¹¹, although by analogy with mice, deleterious pleiotropic effects may be expected with MT-GH constructs¹². Our results, in which substantial amounts of a foreign milk protein were produced from single transgenic loci, show that the protein composition of milk may be manipulated by gene transfer without any apparent adverse consequences. Milk provides 20-30% of the total protein intake of people in the industrialized world¹³. Gene transfer into dairy animals should be viewed as a realistic approach for the production of

Table 1 BLG in milk and BLG mRNA in the mammary gland

Animal	BLG concentration (mg ml ⁻¹)	BLG staining (%)	BLG mRNA
Mouse			
7.1	23.2 ± 0.95	42.7 ± 4.8	3.6
7.10	23.0 ± 1.1	38.8 ± 1.0	2.9
14.10	3.0 ± 0.32	13.8 ± 1.0	0.6
14.11	7.0 ± 0.54	13.3 ± 1.8	0.9
45.18	14.1 ± 0.41	48.6 ± 3.3	3.2
45.20	21.6 ± 1.6	45.5 ± 4.8	3.7
41	8.6 ± 0.88	21.6 ± 0.035	ND
Sheep			
	4.6 ± 1.1	12.6 ± 3.1	1.0

BLG was estimated by densitometry (Joyce-Loebl Chromoscan 3) of Coomassie-Blue-stained SDS-PAGE gels, with reference to internal standards of purified BLG. Over the range of protein concentrations determined, staining intensity was linearly related to the amount of BLG standard. Values for mice are means of determinations from duplicate gels. The value for sheep is the mean of determinations on milk samples from four ewes. The intensity of staining of the BLG band is expressed as a percentage of the total staining of proteins in each milk sample. Values obtained from same gels as BLG concentration. Amounts of BLG mRNA were estimated by dot-blotting 1-µg aliquots of total mammary-gland RNA and probing with p931. The conditions of hybridization were the same as those used for Northern blot analysis. After autoradiography, individual dots were excised and counted in a scintillation counter. Results (average of at least six determinations) are expressed relative to the level of BLG determined by hybridization of p931 to total RNA from the mammary gland of a lactating ewe. ND, not done.

milk with enhanced nutritional value, or with properties more suitable for the dairy industry¹⁴. In addition, it may be possible to engineer dairy animals to produce proteins of high value in their milk^{15,16}. We have produced transgenic sheep carrying DNA segments derived from the BLG gene fused to the sequences encoding human Factor IX and human α1-antitrypsin (J.P.S. *et al.*, unpublished), with a view to collecting these proteins from milk.

We thank R. Lovell-Badge for advice on microinjection, A. Walker and R. Wallace for care of animals, M. Robertson and J. Anderson for technical assistance, J.-C. Mercier for the BLG cDNA clone, P. Gaye for purified BLG and anti-BLG antiserum, J. O. Bishop for advice and use of his facilities, R. Lathe, J. O. Bishop, S. Harris and B. Whitelaw for reading the manuscript, I. Wilmut and R. Al-Shawi for discussions and R. B. Land for support and encouragement.

Received 7 May; accepted 29 June 1987.

- Gaye, P. *et al.* *Biochimie* **68**, 1097-1107 (1986).
- Pervais, S. & Brew, K. *Science* **228**, 335-337 (1985).
- Papiz, M. Z. *et al.* *Nature* **324**, 383-385 (1986).
- Mercier, J.-C., Gaye, P., Soulier, S., Hue-Delahaie, D. & Vilotte, J.-L. *Biochimie* **67**, 959-971 (1985).
- Frischauf, A.-M., Lehrach, H., Poutska, A. & Murray, N. *J. molec. Biol.* **170**, 827-842 (1983).
- Forsyth, I. A. in *The Biochemistry of Lactation* (ed. Mepharm, T. B.) 309-349 (Elsevier, Amsterdam, 1983).
- Forsyth, I. A. *J. Dairy Sci.* **69**, 886-903 (1986).
- Evans, J. W. & Hollaender, A. (eds) *Genetic Engineering of Animals* (Plenum, New York, 1986).
- Simons, J. P. & Land, R. B. *J. Reprod. Fert. suppl.* **34**, 237-250 (1987).
- Smith, C., Meuwissen, T. H. E. & Gibson, J. P. *Animal Breeding Abstr.* **55**, 1-10 (1987).
- Hammer, R. E. *et al.* *Nature* **315**, 680-683 (1985).
- Hammer, R. E., Palmiter, R. D. & Brinster, R. L. *Nature* **311**, 65-67 (1984).
- Hambraeus, L. (1982) in *Developments in Dairy Chemistry-1* (ed. Fox, P. F.) 289-313 (Applied Science, London, 1982).
- Kang, Y., Jimenez-Flores, R. & Richardson, T. (1985) in *Genetic Engineering of Animals* (eds Evans, J. W. & Hollaender, A.) 95-111 (Plenum, New York, 1985).
- Lathe, R. *et al.* in *Exploiting New Technologies in Animal Breeding* (eds Smith, C., King, J. W. B. & McKay, J.) 91-102 (Oxford University Press, 1986).
- Clark, A. J., Simons, J. P., Wilmut, I. & Lathe, R. *Trends Biotechnol.* **5**, 20-24 (1987).
- Flecknell, P. A. *Vet. Rec.* **113**, 574 (1983).
- Laemmli, U. K. *Nature* **227**, 680-685 (1970).
- Khyse-Anderson, J. *J. biochem. biophys. Meth.* **10**, 203-209 (1984).
- Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. *Biochemistry* **18**, 5294-5299 (1979).
- Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).

Ankyrin binding to (Na⁺ + K⁺)ATPase and implications for the organization of membrane domains in polarized cells

W. James Nelson[†] & Pamela J. Veshnock

Institute for Cancer Research, 7701 Burholme Avenue, Philadelphia, Pennsylvania 19111, USA

The interaction between membrane proteins and cytoplasmic structural proteins is thought to be one mechanism for maintaining the spatial order of proteins within functional domains on the plasma membrane¹⁻⁴. Such interactions have been characterized extensively in the human erythrocyte⁵⁻⁷, where a dense, cytoplasmic matrix of proteins comprised mainly of spectrin and actin⁸, is attached through a linker protein, ankyrin, to the anion transporter (Band 3)^{9,10}. In several nonerythroid cell types, including neurons¹¹⁻¹³, exocrine cells¹⁴ and polarized epithelial cells^{15,16} homologues of ankyrin and spectrin (fodrin)^{11,12,15,17-23} are localized in specific membrane domains. Although these results suggest a functional linkage between ankyrin and fodrin and integral membrane proteins in the maintenance of membrane domains in nonerythroid cells, there has been little direct evidence of specific molecular interactions. Using a direct biological and chemical approach, we show here that ankyrin binds to the ubiquitous (Na⁺ + K⁺)ATPase, which has an asymmetrical distribution in polarized cells^{14,24-30}.

We purified the (Na⁺ + K⁺)ATPase in a membrane-bound and active form from canine kidney outer medulla³¹ (Fig. 1a, b). The (Na⁺ + K⁺)ATPase vesicle fraction comprised the α - and β -subunits of the (Na⁺ + K⁺)ATPase^{16,31} (Fig. 1b, lanes 2-4) and minor proteins of relative molecular mass (M_r) 60,000 and 30,000³¹ (Fig. 1b, lane 2). Trypsinization of the (Na⁺ +

K⁺)ATPase vesicles in the presence of 150 mM NaCl produces a characteristic pattern of large protein fragments derived from the α -subunit of the (Na⁺ + K⁺)ATPase (Fig. 1c), showing that the (Na⁺ + K⁺)ATPase is oriented in the membrane with the cytoplasmic domain on the outer face of the vesicles³¹⁻³³. Immunoblotting with specific antisera shows that the (Na⁺ + K⁺)ATPase vesicles are depleted of fodrin and ankyrin (Fig. 1d).

We initially assessed binding of purified ankyrin (Fig. 1e, lane 2) to purified (Na⁺ + K⁺)ATPase vesicles by centrifugation on linear 10-50% (w/v) sucrose density gradients (Fig. 2). Under these conditions, the purified (Na⁺ + K⁺)ATPase vesicles form a single, discrete band in the sucrose gradient (Fig. 2a, b, e), and purified ankyrin remains at the top of the gradients (Fig. 2c). Following incubation with purified (Na⁺ + K⁺)ATPase vesicles, however, ankyrin co-sediments with the (Na⁺ + K⁺)ATPase (Fig. 2d). This result is not apparently caused by trapping of ankyrin in the membrane vesicles, as bound ankyrin is sensitive to proteolysis by trypsin in the absence of detergents (Fig. 2f).

To determine whether ankyrin binds directly to the (Na⁺ + K⁺)ATPase, we sought to isolate such a complex by extracting the (Na⁺ + K⁺)ATPase from the vesicles with the non-ionic detergent octaethyleneglycoldodecyl ether (C₁₂E₈)^{31,34}. C₁₂E₈ solubilizes quantitatively both subunits of the (Na⁺ + K⁺)ATPase (Fig. 3a), which comprise >97% of the protein extracted from the vesicles as determined by scanning densitometry. When the C₁₂E₈-extracted (Na⁺ + K⁺)ATPase subunits were subjected to polyacrylamide gel electrophoresis (PAGE) under non-denaturing conditions (Fig. 3b), we detected a single band (Fig. 3b, panel 1). We showed directly that this band is a complex of α - and β -subunits of the (Na⁺ + K⁺)ATPase by resolving the components on a second-dimension, denaturing SDS-polyacrylamide gel (Fig. 3b, panel 2). Silver staining of this gel reveals only two proteins (M_r 100,000 and 60,000) (data not shown), whose identity as the α - and β -subunits of the (Na⁺ + K⁺)ATPase, respectively, we confirmed immunologically (Fig. 3b, panel 2).

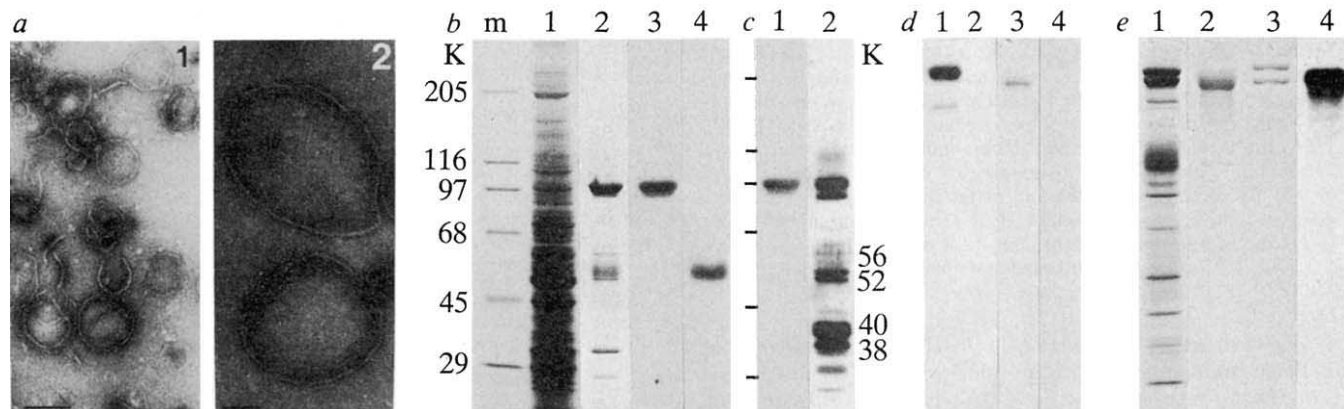


Fig. 1 Characterization of purified (Na⁺ + K⁺)ATPase vesicles and ankyrin. **a**, Electron micrographs of a negatively stained preparation of purified (Na⁺ + K⁺)ATPase vesicles from canine kidney microsomes; bar, 25 μ m (1) and 10 μ m (2). **b**, Coomassie Blue-stained, SDS-PAGE⁴⁹ (lanes m, 1, 2) and corresponding immunoblots¹⁶ (lanes 3, 4) of (Na⁺ + K⁺)ATPase vesicles. Lane m, M_r standards: myosin (250,000 (250K)), β -galactosidase (116K), phosphorylase *b* (97K), bovine serum albumin (BSA) (68K), ovalbumin (45K) and carbonic anhydrase (29K). Lane 1, 250 μ g canine kidney microsomes; lane 2, 24 μ g purified (Na⁺ + K⁺)ATPase vesicles; lanes 3, 4, immunoblots of purified (Na⁺ + K⁺)ATPase vesicles with rabbit polyclonal antiserum specific for the α -subunit (lane 3) or β -subunit (lane 4) of the (Na⁺ + K⁺)ATPase¹⁶. **c**, Characterization by immunoblotting of proteolytic fragments of the α -subunit of the (Na⁺ + K⁺)ATPase following limited digestion of (Na⁺ + K⁺)ATPase vesicles with trypsin in the presence of 150 mM NaCl. Lane 1, undigested α -subunit (Na⁺ + K⁺)ATPase (100K); lane 2, (Na⁺ + K⁺)ATPase vesicles following 10 min incubation with trypsin at 37°C; major, membrane-bound fragments have apparent M_r of 56K, 52K, 40K and 38K. **d**, Identification of fodrin (lanes 1, 2) and ankyrin (lanes 3, 4) in canine kidney microsomes (lanes 1, 3) and purified (Na⁺ + K⁺)ATPase vesicles (lanes 2, 4) by SDS-PAGE and immunoblotting with specific rabbit polyclonal antisera. **e**, Characterization of ankyrin purified from human erythrocytes⁵⁰. Lane 1, human erythrocyte membrane proteins; lane 2, 30 μ g purified ankyrin (M_r ~230K); lane 3, 25 μ g purified spectrin dimers; lane 4, immunoblot of purified ankyrin with a rabbit polyclonal ankyrin antiserum.

Methods. The (Na⁺ + K⁺)ATPase was purified in both forms from canine kidney outer medulla (Pel-Freez Biologicals, Rogers, Arkansas) as described³¹; analysis of ATP hydrolysis activity³⁴ reveals that an average (Na⁺ + K⁺)ATPase preparation releases 75 μ g Pi mg⁻¹ protein min⁻¹ at 25°C. For electron microscopy, (Na⁺ + K⁺)ATPase vesicles were absorbed onto a Formvar grid pre-coated with BSA, negatively stained with 2% (w/v) aqueous uranyl acetate, and viewed in a Philips 420 electron microscope operated at 80 KV. The orientation of the (Na⁺ + K⁺)ATPase in the vesicle membrane was determined by limited proteolysis with trypsin³¹; 48 μ g purified (Na⁺ + K⁺)ATPase vesicles were incubated in a buffer containing 25 mM imidazole, pH 7.2, 1 mM EDTA (buffer A), 150 mM NaCl and 4.8 μ g TPCK-trypsin (Worthington) in a final volume of 50 μ l for 10 min at 37°C, and the reaction terminated by the addition of 15 μ g soybean trypsin inhibitor (Sigma) and centrifugation through 10% (w/v) sucrose in buffer A at 48,000g for 60 min at 4°C. The pellet was analysed by SDS-PAGE and immunoblotting as described.

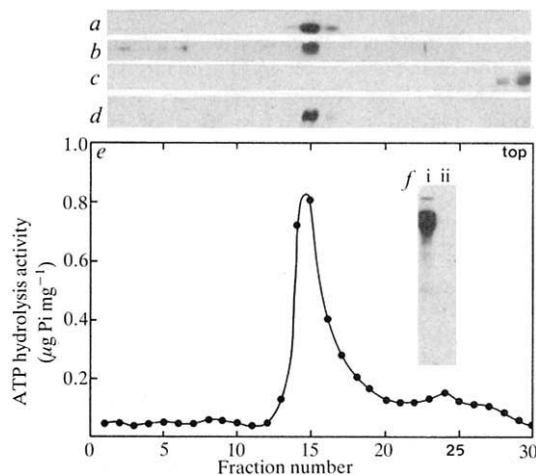


Fig. 2 Sedimentation of purified ($\text{Na}^+ + \text{K}^+$)ATPase vesicles and ankyrin on sucrose density gradients. Purified ($\text{Na}^+ + \text{K}^+$)ATPase vesicles (200–400 μg) in 100 μl buffer A were loaded onto a linear 10%–50% (w/v) sucrose density gradient in 12 ml buffer A and centrifuged at 288,000g for 1 h at 4°C in the SW41Ti rotor of the Beckman L8-70 ultracentrifuge. The gradient was fractionated from the bottom (fraction 1) to the top (fraction 30), and five drops (~400 μl) were collected from each fraction. Each fraction (40 μl) was assayed for ATP hydrolysis activity as described³⁴ (e). The distribution of ($\text{Na}^+ + \text{K}^+$)ATPase subunits on the sucrose gradient was determined by SDS-PAGE and subsequent immunoblotting with α -subunit (a) or β -subunit (b) specific antisera and autoradiography. The relative sedimentation rate of purified ankyrin (50 μg dialysed against binding buffer) was determined in a duplicate sucrose gradient (c). The sedimentation of ankyrin that had been preincubated with purified ($\text{Na}^+ + \text{K}^+$)ATPase vesicles was also analysed on duplicate sucrose gradients as described above, and determined by immunoblotting (d). f, Fraction 15 (100 μl) from a sucrose gradient containing ankyrin and the ($\text{Na}^+ + \text{K}^+$)ATPase (see d) was incubated in the absence (i) or presence (ii) of 50 μg TPCK-trypsin for 15 min at 37°C. The vesicles were centrifuged through 10% (w/v) sucrose in buffer A and analysed by SDS-PAGE and immunoblotting with ankyrin-specific antiserum.

Methods For the binding of ankyrin to the ($\text{Na}^+ + \text{K}^+$)ATPase vesicles, 200–400 μg of the vesicles (80–160 μl in buffer A) were incubated with 2–10 μg ankyrin (50–100 μl , dialysed against binding buffer; 2.5 mM potassium phosphate buffer, pH 7.6, 2 mM KCl 1 mM MgCl_2 , 0.2% (v/v) 2-mercaptoethanol) in a final volume of 250–350 μl in binding buffer containing 1% (w/v) BSA for 90 min at 0°C. In some cases, at the end of the incubation the vesicles were underlayered with 100–200 μl of a buffer containing 10% (w/v) sucrose, 0.7 mM sodium phosphate buffer, pH 7.6, 20 mM KCl, 1% BSA and centrifuged at 48,000g for 1 h at 4°C.

C_{12}E_8 – extraction of purified ($\text{Na}^+ + \text{K}^+$)ATPase vesicles that had been preincubated with purified ankyrin results in the co-solubilization of the ($\text{Na}^+ + \text{K}^+$)ATPase and ankyrin (Fig. 3c). PAGE under nondenaturing conditions reveals that the C_{12}E_8 – extracted ankyrin and ($\text{Na}^+ + \text{K}^+$)ATPase co-migrate as a single, discrete band (Fig. 3d, lane 3), which has an electrophoretic mobility slightly slower than that of the purified ($\text{Na}^+ + \text{K}^+$)ATPase subunit complex (Fig. 3d, lane 2). Immunoblotting with specific antisera reveals the presence of the α -subunit (Fig. 3e, lane 3) and β -subunit (data not shown) of the ($\text{Na}^+ + \text{K}^+$)ATPase, and ankyrin (Fig. 3f, lane 3) in this band. Significantly, little or no protein can be detected by staining with Coomassie blue (Fig. 3d) or by immunoblotting (Fig. 3e, f) with an electrophoretic mobility similar to that of either purified ($\text{Na}^+ + \text{K}^+$)ATPase or ankyrin. Thus, the electrophoretic mobility of all the C_{12}E_8 -solubilized ($\text{Na}^+ + \text{K}^+$)ATPase complex decreases when saturating amounts of ankyrin are bound (see below).

We determined the specificity of binding of ankyrin to ($\text{Na}^+ + \text{K}^+$)ATPase vesicles by the criteria that binding should be

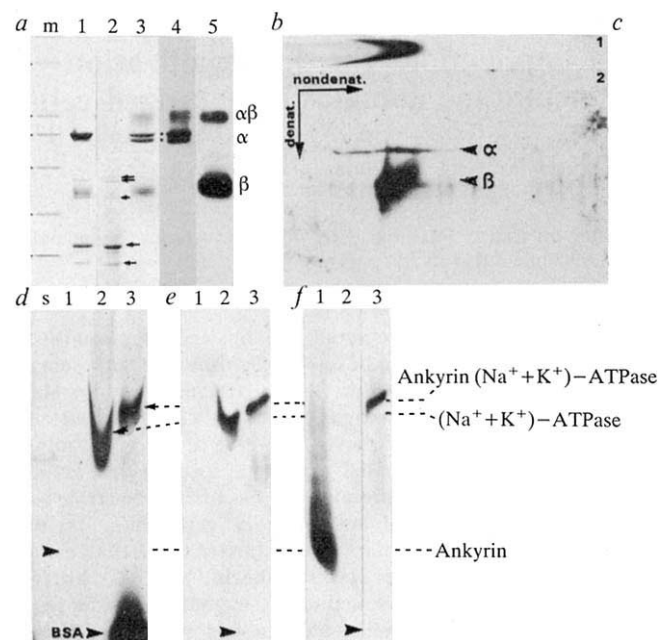


Fig. 3 Characterization by non-denaturing PAGE of an ankyrin-($\text{Na}^+ + \text{K}^+$)ATPase complex solubilized from ($\text{Na}^+ + \text{K}^+$)ATPase vesicles. a, Linear gradient SDS-PAGE (5–12.5%) (lanes m, 1–3) and corresponding immunoblots (lanes 4, 5) of C_{12}E_8 -extracted ($\text{Na}^+ + \text{K}^+$)ATPase vesicles. Lane m, M_r standards (see Fig. 1 legend); lane 1, 35 μg purified ($\text{Na}^+ + \text{K}^+$)ATPase vesicles; lane 2, insoluble residue following C_{12}E_8 extraction of 35 μg purified ($\text{Na}^+ + \text{K}^+$)ATPase vesicles; arrows denote minor contaminating proteins that are insoluble under these extraction conditions; lane 3, proteins extracted from ($\text{Na}^+ + \text{K}^+$)ATPase vesicles with C_{12}E_8 ; lanes 4 and 5, immunoblots of proteins extracted from ($\text{Na}^+ + \text{K}^+$)ATPase vesicles with C_{12}E_8 using antisera specific for either the α - (lane 4) or β -subunits (lane 5) of the ($\text{Na}^+ + \text{K}^+$)ATPase. In some preparations a doublet with an apparent M_r ~100K which reacts with ($\text{Na}^+ + \text{K}^+$)ATPase α -subunit-specific antisera was detected. Undissociated $\alpha\beta$ -subunit complexes with an apparent M_r of 120K were also present (lanes 3–5). b, Characterization by 2–4% non-denaturing (panel 1), and subsequent second-dimension 5–12.5% SDS-denaturing PAGE (panel 2) of ($\text{Na}^+ + \text{K}^+$)ATPase subunits extracted by C_{12}E_8 from ($\text{Na}^+ + \text{K}^+$)ATPase vesicles. Panel 1, Coomassie Blue-stained gel; panel 2, autoradiogram of an immunoblot of the second-dimension denaturing gel using α - and β -subunit-specific ($\text{Na}^+ + \text{K}^+$)ATPase antisera. c, Immunoblot using ankyrin-specific antiserum of the C_{12}E_8 -extracted ($\text{Na}^+ + \text{K}^+$)ATPase vesicle proteins following incubation of ankyrin with ($\text{Na}^+ + \text{K}^+$)ATPase vesicles and subsequent centrifugation through 20% (w/v) sucrose. d, Coomassie Blue-stained non-denaturing 2–4% linear gradient polyacrylamide slab gel. Lane S, 25 μg purified human erythrocyte spectrin dimers (see Fig. 1e, lane 2); lane 1, 4 μg purified ankyrin; lane 2, C_{12}E_8 extract of 24 μg purified ($\text{Na}^+ + \text{K}^+$)ATPase vesicles; lane 3, C_{12}E_8 extract of 24 μg purified ($\text{Na}^+ + \text{K}^+$)ATPase vesicles that had been preincubated with 2 μg purified ankyrin and BSA. e, f, Corresponding immunoblots of panel a, lanes 1–3, using antisera specific for the α -subunit of the ($\text{Na}^+ + \text{K}^+$)ATPase (b) or ankyrin (c). Similar results to those in b were obtained with antisera specific for the β -subunit of the ($\text{Na}^+ + \text{K}^+$)ATPase (data not shown).

Methods. C_{12}E_8 (Calbiochem: 1.68 mg ml^{-1} in buffer B) was added to 24 μg purified ($\text{Na}^+ + \text{K}^+$)ATPase vesicles in a C_{12}E_8 /protein ratio of 5–6³⁴ in binding buffer containing 1% (w/v) BSA. The membranes were incubated with C_{12}E_8 for 5 min at 0°C, and centrifuged at 48,000g for 30 min at 4°C; the resulting soluble and insoluble vesicle material were analysed by SDS-PAGE (denaturing³⁵), or non-denaturing PAGE³². For analysis by non-denaturing and then SDS-denaturing gel, the non-denaturing gel was run in a glass tube (diameter 3 mm) as above. Then the gel was removed from its tube, incubated at room temperature in a buffer containing 1% SDS, 62.5 mM Tris-HCl, pH 6.8, 0.1 mM dithiothreitol for 15 min, then placed on top of a 5–12% linear gradient SDS gel, sealed in place with agarose and electrophoresed at 25 mA constant current. The gel was processed subsequently for silver staining³³ (data not shown), or immunoblotting and autoradiography.

saturable, of high affinity, and susceptible to proteolysis of the vesicles and to heat denaturation of ankyrin. The amount of ankyrin bound to ($\text{Na}^+ + \text{K}^+$)ATPase vesicles is similar in the presence of either Na^+ or K^+ , indicating that changes in the conformation of the membrane-bound ($\text{Na}^+ + \text{K}^+$)ATPase

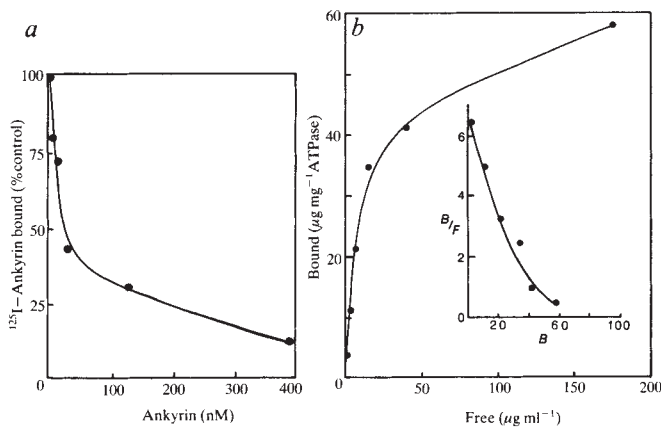


Fig. 4 Characterization of binding of ^{125}I -labelled ankyrin to purified $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles. *a*, Effect of increasing concentration of unlabelled ankyrin on the binding of ^{125}I -labelled ankyrin to purified $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles. *b*, Binding of increasing concentrations of ^{125}I -labelled ankyrin to purified vesicles. Inset, data presented in a Scatchard plot; *b*, ^{125}I -ankyrin bound, $\mu\text{g mg}^{-1}$ protein (vesicles); *F*, free ^{125}I -ankyrin ($\mu\text{g ml}^{-1}$).

Methods. Purified ankyrin was iodinated by the method described previously using Iodogen¹⁶; the specific activity of the ^{125}I -labelled ankyrin was 2.87×10^7 c.p.m. μg^{-1} . For binding analysis, purified $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles (5–25 μg per assay) were incubated in the presence of ^{125}I -labelled ankyrin (1–50 μg) and unlabelled ankyrin (5–600 nM) for 90 min at 4 °C in a final volume of 200 μl binding buffer containing 1% (w/v) BSA (see Fig. 2 legend). Under these conditions >85% of the ^{125}I -ankyrin could bind in excess of $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles (data not shown). Following centrifugation through a layer of 20% (w/v) sucrose (see Fig. 2 legend), the supernatant was discarded and the pellet analysed in a LKB gamma counter (1261 Multigamma) to quantitate bound ankyrin. As controls for the specificity of ankyrin binding, ^{125}I -labelled ankyrin was heat-denatured at 65 °C for 30 min in binding buffer, cooled on ice and then added to $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles as described above. In addition, purified $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles (5 μg) were incubated in the presence of 0.5 μg TPCK-trypsin at 37 °C for 60 min and the reaction stopped by the addition of 2 μg soybean trypsin inhibitor. Under these conditions, all major fragments of the cytoplasmic domain of the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ subunits (see Fig. 1c) were degraded. The vesicles were then incubated with ^{125}I -labelled ankyrin as described above.

induced by these ions³¹ does not affect ankyrin binding (data not shown). However, >90% of the binding site(s) for ankyrin on the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles is susceptible to proteolysis with trypsin, indicating that ^{125}I -labelled ankyrin binds to a protein(s) on the vesicles; in addition, >90% of binding is inhibited if ^{125}I -labelled ankyrin is heat-denatured at 65 °C for 30 min (data not shown; see also Fig. 4 legend).

The capacity of $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles for ankyrin is limited: 2.5 nM ^{125}I -labelled ankyrin could be competed by unlabelled ankyrin, with 50% inhibition occurring at 25–30 nM (Fig. 4a) and >95% inhibition at 800 nM (data not shown). We analysed the affinity of ankyrin for purified $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles by binding increasing concentrations of ^{125}I -labelled ankyrin to a constant concentration of vesicles (Fig. 4b). The Scatchard plot derived from these data is shown in Fig. 4b (inset) and the apparent K_d was estimated as $\sim 1\text{--}2.5 \times 10^{-9}$ M.

These results indicate that the binding of ankyrin to $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles is specific. That ankyrin co-extracts with the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ subunits from these vesicles with C_{12}E_8 in the absence of any apparent contaminating proteins (see Fig. 3) indicates strongly that the ^{125}I -labelled ankyrin binds to the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$. Although the affinity of ^{125}I -labelled ankyrin for $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles is high, the binding is sub-stoichiometric with an extrapolated maximum capacity of 60–

80 μg ankyrin per mg $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ vesicles (Fig. 4b, inset). This compares with a binding capacity of 100–150 μg ankyrin per mg of KI-extracted, inverted red cell membranes, which comprise almost pure anion transporter^{9,10}. Conceivably, the low capacity of the vesicles for ankyrin may be caused by partial denaturation of a proportion of the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ resulting from the treatment with low concentration of SDS during purification. However, note that our analysis of the ankyrin/ $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ complex by non-denaturing PAGE does reveal that binding of substoichiometric amounts of ankyrin alters the electrophoretic mobility of all the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$. This suggests that binding of ankyrin results in the clustering of $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ molecules to form higher complexes, which may not bind further ankyrin molecules because of conformational changes or steric hindrance.

Taken together these results provide evidence of a strong binding between ankyrin, a major component of the membrane-skeleton and the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$. Because ankyrin binds spectrin (fodrin)^{5,6}, to which intermediate filaments³⁵ also bind, and tubulin³⁶, and as the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ has a ubiquitous distribution, our results raise the possibility that this interaction is a common molecular mechanism for linking major components of the cytoskeleton through ankyrin and the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ to the membrane in many eukaryotic cells.

The interaction between ankyrin and $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ also provides an insight into a mechanism for maintaining the organization of polarized cells, in which proteins of different membrane domains must be prevented from mixing^{37–42}. Such constraints to lateral diffusion of proteins could occur passively by physical barriers such as the tight junction delineating the boundary between the apical and basolateral domains of the plasma membrane in polarized epithelial cells^{43–46}. Alternatively, interactions between membrane proteins and the cytoskeleton could actively prevent mixing of membrane proteins through the formation of topographically fixed higher-order structures on the membrane. In polarized kidney epithelial cells *in vivo*¹⁵ and *in vitro*¹⁶, ankyrin, fodrin (spectrin) and the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ are colocalized on the basolateral domain of the plasma membrane. Significantly, the development of a polarized distribution of the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ in cultures of MDCK epithelial cells *in vitro* coincides temporally and spatially with the assembly of the membrane-skeleton on the basolateral domain of the plasma membrane¹⁶. We have suggested previously that one consequence of the assembly of the membrane-skeleton (ankyrin and fodrin) at this location is that membrane proteins targeted to the basolateral domain of the plasma membrane (such as the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ ⁴⁷) become trapped within the membrane-skeleton and hence constrained from diffusion in the plane of the lipid bilayer⁴⁸. The demonstration here of a binding between ankyrin and the $(\text{Na}^+ + \text{K}^+)\text{ATPase}$ provides evidence strongly supporting this hypothesis.

We thank Secretarial Services for typing the manuscript, Sally Shephardson for the electron microscopy, and Dr Harry Maisel, Department of Anatomy, Wayne State University, Detroit, Michigan for the ankyrin antiserum. Supported by grants to W.J.N. from NIH and NSF; from the NIH to the Institute for Cancer Research; and by the Commonwealth of Pennsylvania.

Received 15 May; accepted 11 June 1987.

- Singer, S. J. *Rev. Biochem.* **43**, 805–833 (1974).
- Nicolson, G. L. *Biochim. Biophys. Acta* **457**, 57–108 (1974).
- Edelman, G. M. *Science* **192**, 218–226 (1976).
- Bourguignon, L. Y. W. & Singer, S. J. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5031–5035 (1977).
- Bennett, V. A. *Rev. Biochem.* **54**, 273–304 (1985).
- Branton, D., Cohen, C. M. & Tyler, J. *Cell* **24**, 24–32 (1981).
- Cohen, C. M. *Semin. Hemat.* **20**, 141–158 (1983).
- Lux, S. E., John, K. M. & Karnovsky, M. J. *J. clin. Invest.* **58**, 955–963 (1976).
- Tyler, J. Hargreaves, W. & Branton, D. *J. biol. Chem.* **255**, 7034–7039 (1979).
- Bennett, V. & Stenbuck, P. *J. biol. Chem.* **255**, 2540–2548 (1980).
- Nelson, W. J. & Lazarides, E. *Cell* **39**, 309–320 (1984).
- Riederer, B. M., Zagon, I. S. & Goodman, S. R. *J. Cell Biol.* **102**, 2088–2097 (1986).
- Lazarides, E. & Nelson, W. J. *Science* **220**, 1295–1296 (1983).
- Perrin, D. & Auris, D. *Nature* **315**, 589–592 (1985).
- Drenckhahn, D., Schluter, K., Allen, D. & Bennett, V. *Science* **230**, 1287–1289 (1985).

16. Nelson, W. J. & Veshnock, P. J. *J. Cell Biol.* **103**, 1751-1765 (1986).
17. Bennett, V., Baines, A. J. & Davis, J. Q. *J. cell. Biochem.* **29**, 157-169 (1985).
18. Burridge, K., Kelley, T. & Mangeat, P. *J. Cell. Biol.* **95**, 478-486.
19. Glenney, J. R. & Glenney, P. *Cell* **34**, 503-512 (1983).
20. Bennett, V., Davis, J. & Fowler, W. A. *Nature* **299**, 126-131 (1982).
21. Kay, M., Tracey, C., Goodman, J., Cone, J. & Bassel, P. *Proc. natn. Acad. Sci. U.S.A.* **80**, 6882-6886 (1983).
22. Cox, J. V., Moon, R. T. & Lazarides, E. *J. Cell Biol.* **100**, 1548-1557 (1985).
23. Nelson, W. J. & Lazarides, E. In *Cell Membranes Methods and Reviews* (eds E. Elson, W. Frazier & L. Glaser), **2**, 219-246 (1984).
24. Vorbrodt, A., Konwinski, M., Solter, D. & Koprowski, H. *Dev. Biol.* **55**, 117-134 (1977).
25. Wiley, L. M. & Eglitis, M. A. *Dev. Biol.* **86**, 493-501 (1981).
26. Fleming, T. P. *Bioassays* **4**, 105-109 (1986).
27. Glynn, I. M. & Karlish, S. J. D. *A. Rev. Physiol.* **37**, 13-85 (1975).
28. Hobbs, A. S. & Albers, R. W. A. *Rev. Biophys. Bioengng* **9**, 259-291 (1980).
29. Hesketh, J. E. *Cell Biol. Int. Rep.* **10**, 85-89 (1986).
30. Hootman, S. R. *Int. Rev. Cytol.* **105**, 129-181 (1986).
31. Jørgensen, P. L. *Biochim. Biophys. Acta* **694**, 27-68 (1982).
32. Karlish, S. J. D. & Yates, D. W. *Biochim. Biophys. Acta* **527**, 115-130 (1978).
33. Giotta, G. *J. biol. Chem.* **250**, 5159-5164 (1975).
34. Brothaus, J. R., Møller, J. V. & Jørgensen, P. L. *Biochem. biophys. Res. Commun.* **100**, 146-154 (1981).
35. Georgatas, S. D., & Marchesi, V. T. *J. Cell Biol.* **100**, 1955-1961 (1985).
36. Davis, J. Q. & Bennett, V. *J. biol. Chem.* **259**, 13550-13559 (1984).
37. Matlin, K. S. & Simons, K. *J. Cell Biol.* **99**, 2131-2139 (1984).
38. Misker, D. E., Bard, E. & Rodriguez-Boulant, E. *Cell* **59**, 537-546 (1984).
39. Pfeiffer, S., Fuller, S. D. & Simons, K. *J. Cell Biol.* **101**, 470-476 (1985).
40. Rindler, M. J., Ivanov, I. E., Plesken, H. & Sabatini, D. D. *J. Cell Biol.* **100**, 136-151 (1985).
41. Rodriguez-Boulant, E. *Mod. Cell Biol.* **1**, 119-170 (1983).
42. Simons, K. & Fuller, S. D. *A. Rev. Cell Biol.* **1**, 243-288 (1985).
43. Gonzalez-Mariscal, L., Chavez de Ramirez, B. & Cerejido, M. J. *membr. Biol.* **86**, 113-125 (1985).
44. Pasaur, M. & Ripoch, P. *J. Cell Biol.* **71**, 907-920 (1976).
45. Van Meer, G. & Simons, K. *EMBO J.* **5**, 1455-1464 (1986).
46. Louvard, D. *Proc. natn. Acad. Sci. U.S.A.* **77**, 4132-4136 (1980).
47. Caplan, M. J., Anderson, M. C., Palade, G. E. & Jamieson, J. D. *Cell* **46**, 623-631 (1986).
48. Nelson, W. J. & Veshnock, P. J. *J. Cell Biol.* **104**, 1527-1537 (1987).
49. Laemli, V. K. *Nature* **227**, 680-685 (1970).
50. Bennett, V. *Meth. Enzym.* **96**, 313-324 (1983).
51. Castro, J. & Farley, R. A. *J. biol. Chem.* **254**, 2221-2228 (1979).
52. Morrow, J. S. & Haigh, W. B. *Meth. Enzym.* **96**, 298-304 (1983).
53. Oakley, B. R., Kirsch, D. B. & Morris, N. R. *Analyt. Biochem.* **105**, 361-363 (1980).

Myosin subfragment-1 is sufficient to move actin filaments *in vitro*

Yoko Yano Toyoshima, Stephen J. Kron,
Elizabeth M. McNally*, Kenneth R. Niebling,
Chikashi Toyoshima & James A. Spudich

Department of Cell Biology, Stanford University School of Medicine,
Stanford, California 94305, USA

* Department of Microbiology and Immunology,
Albert Einstein College of Medicine, Bronx,
New York 10461, USA

The rotating crossbridge model for muscle contraction¹ proposes that force is produced by a change in angle of the crossbridge between the overlapping thick and thin filaments. Myosin, the major component of the thick filament, is comprised of two heavy chains and two pairs of light chains. Together they form two globular heads, which give rise to the crossbridge in muscle, and a coiled-coil rod, which forms the shaft of the thick filament. The isolated head fragment, subfragment-1 (S1), contains the ATPase and actin-binding activities of myosin (Fig. 1). Although S1 seems to have the requisite enzymatic activity, direct evidence that S1 is sufficient to drive actin movement has been lacking. It has long been recognized that *in vitro* movement assays are an important approach for identifying the elements in muscle responsible for force generation²⁻⁸. Hynes *et al.*⁷ showed that beads coated with heavy meromyosin (HMM), a soluble proteolytic fragment of myosin consisting of a part of the rod and the two heads, can move on *Nitella* actin filaments³. Using the myosin-coated surface assay of Kron and Spudich⁶, Harada *et al.*⁸ showed that single-headed myosin filaments bound to glass support movement of actin at nearly the same speed as intact myosin filaments. These studies show that the terminal portion of the rod and the two-headed nature of myosin are not required for movement. To restrict the region responsible for movement further, we have modified the

myosin-coated surface assay by replacing the glass surface with a nitrocellulose film. Here we report that myosin filaments, soluble myosin, HMM or S1, when bound to a nitrocellulose film, support actin sliding movement (Fig. 2). That S1 is sufficient to cause sliding movement of actin filaments *in vitro* gives strong support to models of contraction that place the site of active movement in muscle within the myosin head.

When skeletal muscle myosin was assembled into filaments and bound to a nitrocellulose surface, fluorescently labelled actin filaments in the presence of ATP moved over the myosin-coated surface at a similar rate to that of actin movement on myosin filaments bound to glass⁶ (Fig. 3). When soluble myosin was applied to a nitrocellulose surface, actin filaments moved over the individual myosin molecules at a similar rate to that observed for myosin filaments, suggesting that the rate of movement on myosin bound to a surface is relatively independent of the organization of the myosin molecules.

The rate of actin sliding movement on a nitrocellulose surface coated with HMM was higher than that for myosin filaments (Fig. 3). Chymotryptic HMM⁹, purified by high performance gel filtration (Fig. 4), was applied at a concentration of 20 $\mu\text{g ml}^{-1}$ to a nitrocellulose surface. In the absence of ATP, actin filaments attached to the HMM-coated surface, immobilized along their full lengths. When ATP was added, nearly all the filaments that had associated with the surface could be observed to move smoothly in winding paths for hundreds of micrometres.

A greater percentage of the bound actin filaments moved on an HMM-coated surface than on myosin filaments bound to glass⁶ (>99% versus ~70%). Movement on an HMM-coated surface depended strongly on the concentration of HMM applied to the nitrocellulose. When HMM was used at concentrations below 5 $\mu\text{g ml}^{-1}$, the density of the bound HMM was not sufficient to support actin movement. Few actin filaments attached to the surface in the absence of ATP and, on addition of ATP, the bound actin filaments detached from the surface. When the concentration applied was increased above 5 $\mu\text{g ml}^{-1}$, movement of actin filaments was observed and the moving filaments became progressively more fragmented as the concentration of HMM was increased.

In preliminary experiments, we examined the density of HMM bound to nitrocellulose when applied at 20 $\mu\text{g ml}^{-1}$. The K-EDTA ATPase activity of an HMM-coated surface was that expected from a density of 500 HMM molecules per μm^2 , which corresponds to an average nearest-neighbour distance of 45 nm. This estimate indicates that, in this assay, the density of myosin heads available to interact with actin filaments is of the order of the density of heads that can interact with thin filaments in muscle.

We prepared three forms of S1 that differ in heavy-chain length¹⁰ and composition of light chains (LC)^{11,12}. Myosin filaments were digested either with papain in the presence or absence of divalent cation to generate papain-Mg.S1 (P-Mg.S1) or papain EDTA.S1 (P-EDTA.S1), or with α -chymotrypsin to prepare chymotryptic S1 (CT-S1). The relative molecular masses of the three forms of S1 are reported to be 130,000 (130K) (P-Mg.S1) and 110K (P-EDTA.S1 and CT-S1)¹³. Each form of S1 was purified by high performance gel filtration (Fig. 4) and assayed for motility to define further the myosin domains required for actin sliding movement.

P-Mg.S1 consists of a 96K amino-terminal fragment of the myosin heavy chain, LC2 and either LC1 or LC3¹¹. When P-Mg.S1 was applied to nitrocellulose surfaces at 20 $\mu\text{g ml}^{-1}$, it bound actin filaments in rigor in the absence of ATP. When ATP was added, nearly all the bound actin filaments began to move. Like HMM or myosin, as the applied concentration of S1 was raised, increasing fragmentation of the filaments on the surface is observed. The average speed of actin movement directed by P-Mg.S1 was slower than that supported by myosin or HMM (Fig. 3).

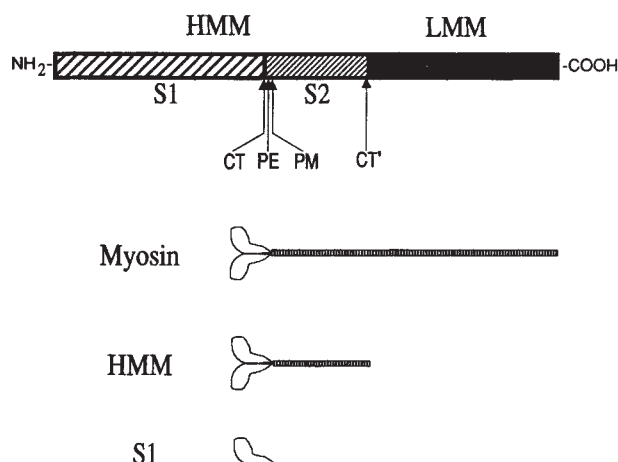


Fig. 1 Diagrammatic representation of the skeletal muscle myosin heavy chain and the soluble myosin fragments used in this study. A coiled-coil rod is formed from the carboxy-terminal halves of the two heavy chains, and the amino termini fold to form individual globular heads. There are two regions of the heavy-chain sequence particularly susceptible to proteolysis. The first is at the 'hinge' region in the rod, where chymotrypsin (CT) cuts to form heavy meromyosin (HMM) and light meromyosin (LMM). HMM is a two-headed soluble fragment with the enzymatic activities of myosin and LMM has the assembly properties of myosin. The second proteolytic site is at the S1-S2 junction, where the heads join the rod. Proteolysis here, by papain with or without divalent cation (PM, PE) or by chymotrypsin (CT), frees the soluble head fragment, subfragment-1 (S1). S1 contains the ATPase, the actin binding and, as shown in this report, the movement activities of the parent molecule, myosin.

P-EDTA.S1 was a mixture of at least two different heavy-chain fragments with mobilities of 95 and 92K on SDS-PAGE (Fig. 4). These two species eluted together at 145K from a gel-filtration column. The LC1 and LC3 in the P-EDTA.S1 were intact, but most of the LC2 was degraded. The degradation of LC2 did not affect the rate of actin sliding movement, as the average speed directed by P-EDTA.S1 was almost the same as that of P-Mg.S1 (Fig. 3).

CT-S1, when bound to nitrocellulose, could also produce actin sliding movement. Chymotryptic digestion resulted in the loss of LC2 and produced heavy chains of 93 and 90K. High-performance gel filtration resolved a 93K-rich fraction which directed actin movement in the assay when diluted to $40 \mu\text{g ml}^{-1}$ and a 90K-rich fraction which supported movement only at concentrations higher than $200 \mu\text{g ml}^{-1}$. The two fractions had equivalent actin-activated ATPase activity. The rate of movement directed by $40 \mu\text{g ml}^{-1}$ CT-S1 enriched with the 93K heavy chain was about half that of the papain S1 species (Fig. 3).

We have two lines of evidence that minor levels of contaminating species cannot explain the S1-directed movement of actin. First, we used high-performance gel filtration to resolve the supernatant of a papain-Mg²⁺ digestion of myosin filaments into two movement activities. The first movement activity, the contaminating HMM and myosin, eluted near the void volume. The concentration of HMM in the trailing edge of this peak was not sufficient to support movement. The second activity, which eluted as a peak at 145K, was P-Mg.S1 and was essentially homogeneous as determined by SDS-PAGE (Fig. 4). When rechromatographed, the P-Mg.S1 eluted as a single peak which directed movement of actin filaments. Second, in titration experiments, when $<5 \mu\text{g ml}^{-1}$ HMM was applied to nitrocellulose, actin filaments did not move on the surface, whereas $20 \mu\text{g ml}^{-1}$ P-Mg.S1 was more than enough to support actin movement. These data suggest that even a contamination of 10% myosin or HMM, much higher than the contamination seen on SDS-PAGE of each S1 preparation, could not have caused the observed movement.

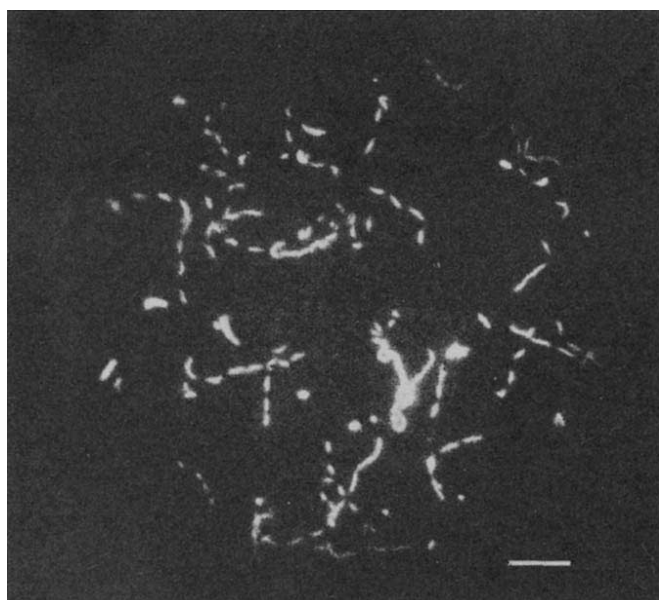


Fig. 2 Stroboscopic photograph of the movement of fluorescent actin filaments on papain Mg.S1. A 35-mm camera was focused on the phosphor screen of a Ni-Tec NVS-100 (Optoelectronic Corp.) image intensifier receiving the light from a $16\times$ eyepiece in the phototube of a Zeiss standard microscope equipped with epi-fluorescence optics and a $63\times$ Planapochromat objective. A photographic shutter was used to illuminate the filaments for 1 s every 4 s for 12 s. The successive images of the actin filaments trace the path of movement. Scale bar, $10 \mu\text{m}$.

Methods. As ref. 6 except that all experiments were carried out in flow cells that had coverslips coated with nitrocellulose films, prepared by spreading 0.5% nitrocellulose (Fullam) onto distilled water. Total volume of the flow cell, $30 \mu\text{l}$. Assay conditions: skeletal muscle myosin fragments or filaments in movement assay buffer (25 mM imidazole, pH 7.4, 25 mM KCl, 4 mM MgCl₂, 1 mM EGTA and 1 mM dithiothreitol, (DTT)) were introduced to the flow cell. After 1 min, the unbound material was washed out of the flow cell with two volumes of 0.5 mg ml^{-1} bovine serum albumin in assay buffer. In the case of soluble myosin, the protein was applied to the flow cell in assay buffer with 0.6 M KCl and unbound material was washed from the flow cell with assay buffer with 0.6 M KCl. Then, one volume of a 1 to 75 dilution of $0.78 \mu\text{M}$ fluorescently labelled actin was infused into the flow cell. Rabbit skeletal muscle actin, prepared by the method of Spudich and Watt²⁰ and recycled²¹, was labelled with tetramethylrhodamine-phalloidin (Molecular Probes) as described⁶. Typically, the actin filaments were allowed to bind to the surface in the absence of ATP, then two volumes of assay buffer with 1 mM ATP was washed through the flow cell, and the movement of actin was observed as described⁶. All experiments were carried out at 30°C using a temperature-controlled slide carrier and temperature-jacketed objective.

The *in vitro* rate of movement directed by S1 was $1\text{--}2 \mu\text{m s}^{-1}$, and the rate of movement directed by HMM was faster, $7 \mu\text{m s}^{-1}$. In contrast, the maximum velocity, v_{max} , for the actin-activated ATPase activities of these preparations of S1 and HMM were very similar (23 s^{-1} for P-Mg.S1; 13 s^{-1} for P-EDTA.S1; 24 s^{-1} for CT-S1; 20 s^{-1} per head for HMM). Therefore, the differences in the rates of movement directed by S1 and HMM in this assay cannot be explained by differences in the turnover rates of ATP.

One possible explanation for the faster rate of movement of HMM than S1 is that a double-headed myosin fragment is inherently better at producing movement than a single-headed fragment. In preliminary experiments, we prepared nearly homogeneous single-headed HMM which supported the sliding movement of actin filaments at speeds higher than those produced by HMM or myosin. This result suggests that the single-headed nature of S1 is not responsible for its slow movement.

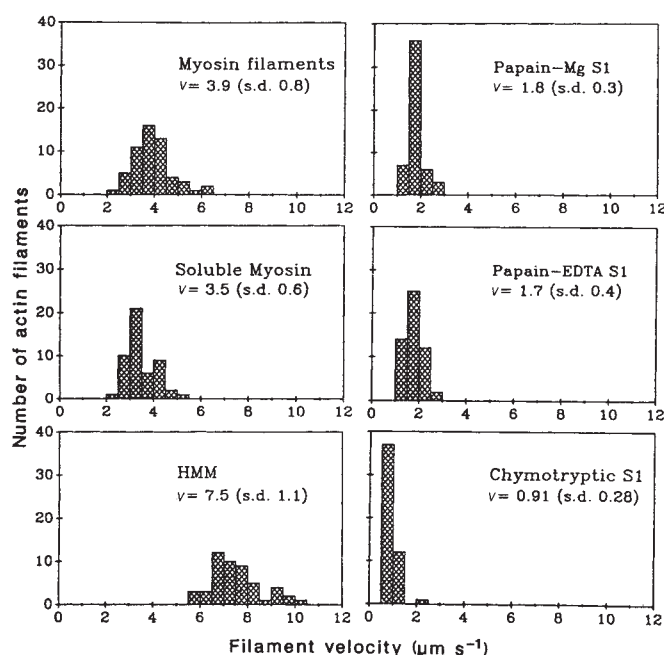


Fig. 3 Histograms comparing the sliding velocities of actin filaments over skeletal muscle myosin and the different soluble fragments. The mean velocity (v) and standard deviation (s.d.) are indicated for each preparation. Speeds were measured by noting the time and position of the beginning and end of smooth, straight paths $>5\mu\text{m}$ in length. Each histogram shows the sliding speed of ≥ 50 individual actin filaments moving on a nitrocellulose surface coated with the indicated form of myosin or its fragments in a single experiment. The indicated mean speeds are characteristic of each form of myosin and reproducible in multiple experiments. The two forms of myosin (upper two panels) directed movement at a rate which was similar to that measured for movement on native skeletal muscle thick filaments on glass ($4\mu\text{m s}^{-1}$, our unpublished results) as well as that measured for movement on synthetic thick filaments on glass⁶. In comparison to myosin, HMM (third panel) produces faster movement while the three forms of S1 (bottom three panels) produces somewhat slower movement. Concentrations of myosin or its fragments applied to the flow cells for these experiments: myosin filaments, $30\mu\text{g ml}^{-1}$; soluble myosin, $20\mu\text{g ml}^{-1}$; HMM, $20\mu\text{g ml}^{-1}$; P.Mg-S1, $40\mu\text{g ml}^{-1}$; P.EDTA-S1, $30\mu\text{g ml}^{-1}$; and CT-S1, $40\mu\text{g ml}^{-1}$.

The different rates of movement on S1 and HMM could be explained by a difference in attachment of these species to the nitrocellulose substratum. HMM may bind to the nitrocellulose via its S2 portion, allowing a productive interaction of many of the heads with actin. S1 may not bind to nitrocellulose with such a favourable orientation; some of the S1 can bind to actin but may not be able to move, causing a restraining force. The resulting balance of forces could result in the relatively slower rate of movement of S1.

The productive interaction of actin and myosin is considered to be the result of a specific binding of the myosin head to actin. In this assay, continuous, directional actin movement occurs despite a random orientation of the myosin fragments on the surface, which implies that there must be torsional freedom within the actin-myosin system. Actin filaments are known to have considerable flexibility¹⁴⁻¹⁶. Flexibility about the S1/S2 junction allows freedom of rotation of the myosin heads^{17,18}, and HMM-directed movement might be facilitated by rotational freedom at this site. That S1 can direct actin sliding movement while it is bound to a solid support may suggest that torsional freedom is present within S1.

S1, actin and ATP constitute a model system for the study of the enzymatic activity responsible for muscle contraction. We

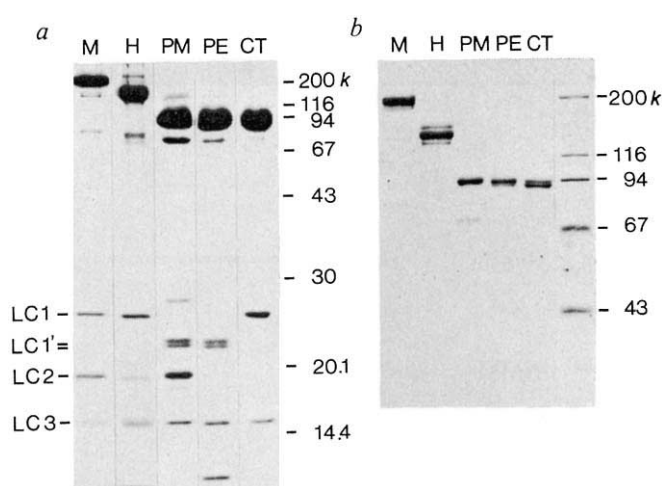


Fig. 4 SDS-PAGE²² of the myosin fragments. *a*, Overloaded ($>15\mu\text{g}$ per lane) 12.5% acrylamide gel, demonstrating the purity and the light chain content of myosin (M) and the soluble fragments (H, HMM; PM, P-Mg.S1; PE, P-EDTA.S1; CT, CT-S1). Both chymotrypsin and papain-EDTA digestions resulted in the degradation of LC2 (DTNB light chain) (lanes H, PE and CT). Papain digestion cleaved LC1 into two fragments of similar mobility (LC1') (PM and PE). Digestion of myosin to S1 resulted in partial cleavage of the heavy chain into 25K and 70K subfragments, most easily seen in lane PM. *b*, An 8% acrylamide gel demonstrating the relative molecular masses of the heavy chains of the different S1 preparations: P-Mg.S1 has a single 96K heavy chain; P-EDTA.S1 has two heavy chains of 95 and 92K; CT-S1 has two heavy chains of 93 and 90K. The masses indicated show the mobility of electrophoresis standards (Pharmacia) with β -galactosidase and skeletal muscle myosin heavy chain added).

Methods. Rabbit skeletal muscle myosin was prepared as described⁴. Chymotryptic HMM was prepared as described by Okamoto and Sekine⁹ except that the digestion buffer was 10 mM imidazole, pH 7.0, 0.5 M KCl, 2 mM MgCl_2 and 1 mM DTT, digestion time 10 min. P-Mg.S1 and P-EDTA.S1 were prepared by digestion of myosin filaments at 12 mg ml^{-1} in 10 mM TES pH 7.0, 0.1 M KCl, 2 mM MgCl_2 or 2 mM EDTA, and 1 mM DTT. Digestion was for 8 min at 25°C with $8\mu\text{g ml}^{-1}$ papain (Sigma, Type IV) and the reaction was stopped with $0.16\mu\text{g ml}^{-1}$ E-64²³ (Peptide Institute). Chymotryptic S1 was prepared as ref. 9 except that digestion was for 3 min. After all digestions, the insoluble material was pelleted. The supernatants were concentrated using a Centricon-30 ultrafiltration device (Amicon) and then loaded onto a Superose 12 gel filtration column (Pharmacia). The appropriate peaks were pooled and assayed for movement and ATPase activity using $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ ²⁴. Reagents and enzymes from Sigma. Protein concentrations were determined by absorbance^{25,26}.

have shown here that this simple biochemical system can produce the sliding movement of actin filaments *in vitro*, and at a velocity that is similar to the maximum shortening velocity of glycerinated muscle under zero load ($6\mu\text{m s}^{-1}$)¹⁹. By demonstrating that S1 can produce actin sliding, we now prove that no portion of the myosin rod is required for movement. This result suggests that the movement which results in muscle contraction occurs in the myosin head.

We have established an *in vitro* motility assay for the soluble fragments of myosin which is both simple and highly reproducible. With this assay, physiologically relevant properties of the actin-myosin interaction can be investigated. By measuring the ATPase associated with the movement in the assay, the step size of head movement for an ATP hydrolysis cycle and the minimum number of heads required to move an actin filament may be determined.

We thank Dr K. Yamamoto for the gift of E-64. Supported by a grant from NIH to J.A.S.; Y.Y.T. is a visiting scientist supported by the Japanese Ministry of Education, Science and

Culture; S.J.K. and E.M.M. are trainees of the Medical Scientist Training Program (NIH); C.T. is a Postdoctoral Fellow of the Muscular Dystrophy Association.

Received 28 April; accepted 17 June 1987.

- Huxley, H. E. *Science* **164**, 1356–1366 (1969).
- Yano, M., Yamamoto, Y. & Shimizu, H. *Nature* **299**, 557–559 (1982).
- Sheetz, M. P. & Spudich, J. A. *Nature* **303**, 31–35 (1983).
- Higashi-Fujime, S. *J. Cell Biol.* **101**, 2335–2344 (1985).
- Honda, H., Nagashima, H. & Asakura, S. *J. molec. Biol.* **191**, 131–133 (1986).
- Kron, S. J. & Spudich, J. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6272–6276 (1986).
- Hynes, T. R., Block, S. M., White, B. T. & Spudich, J. A. *Cell* **48**, 953–963 (1987).
- Harada, Y., Noguchi, A., Kishino, A. & Yanagida, T. *Nature* **326**, 805–808 (1987).
- Okamoto, Y. & Sekine, T. *J. Biochem., Tokyo* **98**, 1143–1145 (1985).
- Yamamoto, K. & Sekine, T. *J. Biochem., Tokyo* **87**, 219–226 (1980).
- Margossian, S. S., Lowey, S. & Barshop, B. *Nature* **258**, 163–166 (1975).
- Yagi, K. & Otani, F. *J. Biochem., Tokyo* **76**, 365–373 (1974).
- Margossian, S. S., Stafford III, W. F. & Lowey, S. *Biochemistry* **20**, 2151–2155 (1981).
- Thomas, D. D., Seidel, J. C. & Gergely, J. *J. molec. Biol.* **132**, 257–273 (1979).
- Egelman, E. H., Francis, N. & DeRosier, D. J. *Nature* **298**, 131–135 (1982).
- Yanagida, T., Nakase, M., Nishiyama, K. & Oosawa, F. *Nature* **307**, 58–60 (1984).
- Mendelson, R. A., Morales, M. F. & Botts, J. *Biochemistry* **12**, 2250–2255 (1973).
- Thomas, D. D., Seidel, J. C., Hyde, J. S. & Gergely, J. *Proc. natn. Acad. Sci. U.S.A.* **72**, 1729–1733 (1975).
- Crowder, M. S. & Cooke, R. *J. Muscle Res. Cell Motil.* **5**, 131–146 (1984).
- Spudich, J. A. & Watt, S. *J. biol. Chem.* **246**, 4866–4871 (1971).
- Pardee, J. D. & Spudich, J. A. *Meth. Cell Biol.* **24**, 271–289 (1982).
- Laemmli, U. K. *Nature* **227**, 680–685 (1970).
- Hanada, K. *et al. Agric. biol. Chem.* **42**, 523–528 (1978).
- Clarke, M. & Spudich, J. A. *J. molec. Biol.* **86**, 209–222 (1974).
- Margossian, S. S. & Lowey, S. *Meth. Enzym.* **85**, 55–71 (1982).
- Gordon, D. J., Yang, Y.-Z. & Korn, E. D. *J. biol. Chem.* **251**, 7474–7479 (1976).

Sequence of simian immunodeficiency virus and its relationship to the human immunodeficiency viruses

G. Franchini†, C. Gurgot, H.-G. Guot, R. C. Gallo†, E. Collalti†§, K. A. Fargnoli*, L. F. Hall†, F. Wong-Staal† & M. S. Reitz Jr†

* Biotech Research Laboratory, Inc., Three Taft Court, Rockville, Maryland 20850, USA

† Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

‡ Centro di Endocrinologia ed Oncologia Sperimentale del CNR Napoli, Italy

The characterization of HIV-1 (HTLV-III/LAV)^{1,2}, the human retrovirus associated with AIDS (acquired immune deficiency syndrome)³ has led to the identification of a group of related human and simian retroviruses which also infect CD4-bearing T lymphocytes. Simian T-lymphotropic virus type III (simian immunodeficiency virus) from macaques (STLV-III_{MAC}) induces symptoms similar to those of AIDS in infected macaques^{4,5}, but isolates from African green monkeys (STLV-III_{AGM})⁶ and mangabeys (STLV-II_{MM})^{7,8} appear to be non-pathogenic in these animals. A human virus immunologically related to STLV-III_{AGM} (HTLV-IV), reported to have been isolated from healthy humans⁹, has been shown to be almost identical to STLV-III_{AGM}^{10,11}, which has called into question the independent origin of these viruses¹¹. Here we report the complete DNA sequence of STLV-III_{AGM} and analyse its relationship with the genomes of the HTLV-III_B strain of HIV-1, HIV-2_{ROD} (previously called LAV-2)^{12–15} and several ungulate lentiretroviruses. STLV-III_{AGM} and HIV-2 are closely related, and more distantly related to HIV-1.

The sequence of the STLV-III_{AGM} provirus, derived by combining the DNA sequences of five overlapping λ clones, is presented in Fig. 1. Comparison of this sequence with those of HTLV-IV (B. Hahn and G. Shaw, personal communication) cDNA clones, of STLV-III_{AGM} from other cell lines, and HIV-

2_{ROD}¹⁵ indicated that our sequence lacks 350 nucleotides in the middle region. The deletion includes part of the *X* region, most of the *R* gene and the first 14 amino acids of the *tat3* gene protein (Fig. 2). Hybridization of an oligomeric probe for this deleted region to the cellular DNA from which these clones were derived indicated that these sequences were also absent in the majority of the proviruses in the infected cells. The homology between STLV-III and HIV-1 at the nucleotide sequence level is 55% overall.

The STLV-III_{AGM} long terminal repeat (LTR) is 800 base pairs long. To define the boundaries of the LTRs, we compared the junction sequence of the *U3*-cellular sequence of the 5' LTR and the *U3*-3'-orf sequence at the 3' end of the virus. Similarly, the end of *U5* was identified by comparing the sequence of the 3' viral-cellular junction and the 5' LTR-gag junction. The boundaries of the *R* region with respect to *U3* and *U5* were determined by comparison with HIV-1^{16–19} and with the sequence of several STLV-III_{AGM} cDNA clones (Colombini-Hatch *et al.*, personal communication). The sizes of the *U3*, *R* and *U5* regions are 498, 176 and 126 base pairs respectively. The sequence TATAA, found 25 nucleotides upstream from the start of transcription in many eukaryotic genes²⁰, is located in *U3* at position –25 from the presumed boundary of the *U3* and *R* regions. The consensus sequence, AATAAA, which signals the addition of the polyadenylate tail²¹, is located at position 152 to position 157.

The primer binding site (PBS), located downstream from the *U5* of the 5' LTR, is complementary to the same isoaccepting form of transfer RNA (tRNA_{lys3}) as are those of HTLV-III_B and other HIV-1 strains^{16–19}. Surprisingly, in view of the close relatedness of these viruses, the STLV-III_{AGM} LTR is much longer than the HIV-1 LTR^{16–19} (800 versus 634 base pairs). The greater size of the STLV-III_{AGM} LTR as well as the HIV-2 LTR¹⁴ relative to that of HIV-1 is due to the presence of several regions in STLV-III_{AGM} of sequences not found in the HIV-1 LTR. A detailed analysis of the STLV-III_{AGM} LTR and the LTRs of other known retroviruses is reported elsewhere²².

The first large open reading frame in the STLV-III_{AGM} genome starts at nucleotide 539 (Fig. 1) and could encode 506 amino acids. The length of the *gag* precursor polypeptide reported for different HIV-1 strains varies between 500 and 512 amino acids. Alignment of STLV-III_{AGM} and HTLV-III_B amino-acid sequences suggests that the p17 homologue contains 135 amino acids and is probably generated by cleavage between Tyr and Pro at position 941–946. Proline is the first amino acid at the N-terminus of HIV-1 p24,^{16–19} the p26 of the ungulate lentiviruses EIAV and Visna^{23,24} the HTLV-1 p24,²⁵ and the HIV-2_{ROD} p26¹⁵. The overall homology of the *gag* precursors of STLV-III_{AGM} to those of HTLV-III_B and HIV-2_{ROD} is 51% and 82% respectively (see Table 1). The highest amino-acid identity can be found between the major core proteins of STLV-III_{AGM} and HTLV-III_B (66%) and HIV-2_{ROD} (88%), consistent with their strong immunological cross-reactivity^{10,13}. The major core protein of the lentivirus equine infectious anaemia virus (EIAV) has slightly higher homology to STLV-III_{AGM} (29%) and HIV-1 (29%), than that of the lentivirus visna (26% and 25%, respectively).

The HTLV-III_B *gag* protein precursor contains two repeated sequences at the carboxy terminus¹⁶. The first, an imperfect repeat encoding 12 amino acids, is also present in STLV-III_{AGM} (see Fig. 1), and in HIV-2_{ROD} at positions 1,605–1,641 and 1,667–1,703, suggesting that this duplication occurred in a common, distant ancestor of these three viruses. The second found in HTLV-III_B is a perfect repeat encoding 12 amino acids¹⁶, which is absent in the STLV-III_{AGM} genome and in some other HIV-1 isolates²⁶ suggesting relatively recent duplication.

The second large open reading frame in the STLV-III_{AGM} genome, nucleotides, 1,714–4,875, has a coding potential for 1,053 amino acids and overlaps the *gag* gene by 342 bases (Fig. 1). We assume that either a splicing or a frameshift event allows

§ Permanent address: Dipartimento Di Medicina Sperimentale, II Cattedra Patologia Generale, Facoltà Di Medicina E Chirurgia, Università Degli Studi, La Sapienza, Rome, Italy.

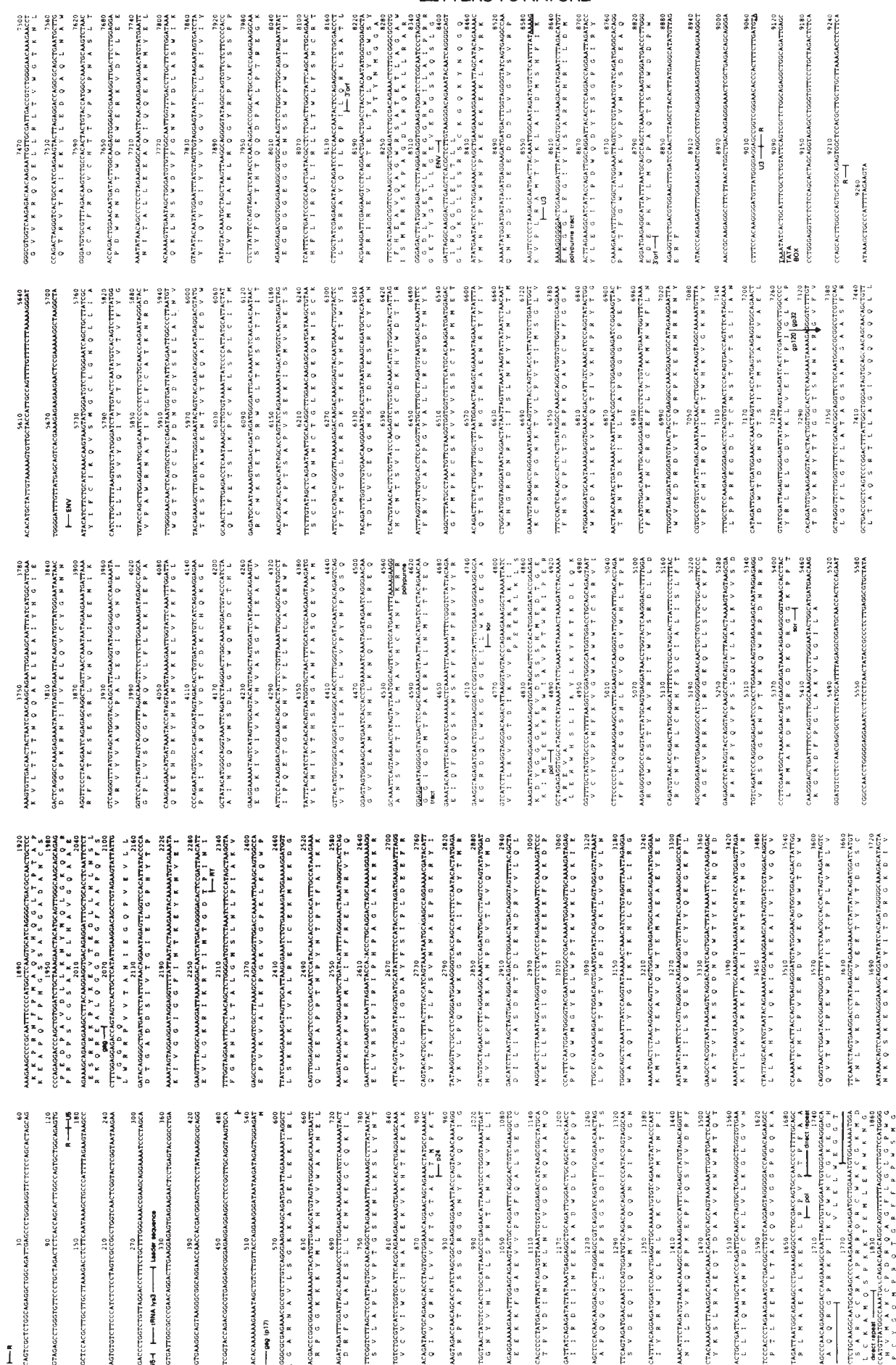
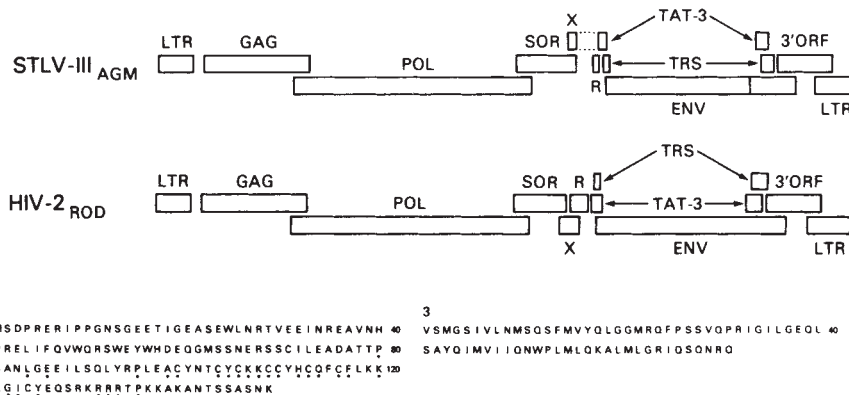


Fig. 1 Sequence of the STLV-III_{AGM} provirus was obtained with the dideoxy chain termination method³⁷ (on single- and double-stranded DNA). The sequence was obtained from both strands of the proviral DNA fragments. The DNA sequence has been numbered from the beginning of the R region at the 5' end of the R region at the 3' end of the proviral DNA. The sequence contains a deletion in the middle part of the genome (see also Fig. 2), here denoted by an arrow (nucleotide 5,468).

Fig. 2 Diagrammatic representation of the genomic structure of STL_V-III_{AGM} and HIV-2_{ROD}¹⁵ proviruses. In the lower panel, the open reading frames identified in the middle portion of the genome are given in single-letter code. 2, 3, second and third open reading frames, respectively. The deletion in STL_V-III_{AGM} found by comparison with HTLV-IV and HIV-2_{ROD} is shown.

Methods. The proviral DNA was obtained from a genomic library constructed from the DNA of the infected cell line K6W using the lambda phage vector EMBL-3. Nine clones containing various overlapping portions of the STL_V-III_{AGM} genome were purified and characterized by restriction endonuclease analysis. Five of them were used to generate viral DNA fragments for subcloning in appropriate vectors.



translation of *pol*. By analogy with HIV-1 and other retroviruses²⁴, this open reading frame should contain the genetic information for the protease, the reverse transcriptase and the endonuclease. In HTLV-III_B, the *pol* region encodes 1,003 amino acids and also extensively overlaps the *gag* reading frame. The amino terminus of the HTLV-III_B reverse transcriptase, beginning with a Pro-Ile, has been identified by amino acid sequencing of the purified p66/p51 viral enzyme²⁷. The amino termini of the HTLV-III_B and HIV-2_{ROD} reverse transcriptase and the inferred amino acid sequence of the STL_V-III_{AGM} RT gene align perfectly. Thus, it is likely that the Pro-Ile in position 2,326-2,331 represents the first two amino acids of the STL_V-III_{AGM} reverse transcriptase. The overall amino-acid homology of the reverse transcriptase of STL_V-III_{AGM} versus HTLV-III_B and HIV-2_{ROD} is 53% and 76% respectively (see Table 1). The amino-acid sequence upstream from the beginning of the reverse transcriptase is also well conserved between HTLV-III_B and STL_V-III_{AGM} and is probably the protease domain. The protease is encoded in the same open reading frame as reverse transcriptase²⁴. The high degree of relatedness of the STL_V-III_{AGM} amino acid sequence in this region with that of HIV-2_{ROD} (88%) HTLV-III_B (49%), EIAV (37%) and visna virus (30%) allows a more precise localization of the presumed cleavage site (Leu-Pro) between the protease and the reverse transcriptase.

We identified three open reading frames in the middle portion of the STL_V-III_{AGM} genome. The first, *sor*, (also referred to as Q), overlaps by 97 nucleotides with the end of the *pol* gene with a potential start codon 12 amino acids from the beginning and could encode a 225-amino acid protein. This protein²⁸ shares 24% amino acid homology with the *sor* gene of HTLV-III_B¹⁶ and other HIV-1 strains¹⁷⁻¹⁹, which is surprisingly low because *sor* is well conserved among various HIV-1 isolates. The amino acid hydropathy profile of the simian and human viral proteins is very similar, however (data not shown). Higher homology can be found between the *sor* (Q) genes of HIV-2_{ROD} and STL_V-III_{AGM} (64%).

The second short open reading frame (position 5,277-5,723) could encode 149 amino acids (see Fig. 2). Its first 60 amino acids share high homology with the X open reading frame described in the HIV-2_{ROD} genome¹⁵. The last part of the sequence shares considerable homology (40%) with the second exon of the transactivator gene (*tat3*) of HIV-1²⁹ which includes the functionally important domains of the *tat3* protein product. The amino acids conserved between the STL_V-III_{AGM} *tat3* and the HTLV-III_B *tat3* are indicated by asterisks in the lower part of Fig. 2. This co-frame fusion of X and *tat* is brought about by the small deletion shown in Fig. 2.

The third open reading frame (5774-5987) encodes 71 amino acids and has a potential initiating methionine in position 3 (Fig. 2).

The largest open reading frame in the 3' region of the STL_V-III_{AGM} genome probably encodes the envelope protein. The first residue is at position 5,702 and the probable initiating methionine codon is located 10 amino acids downstream. The *env* gene probably extends to a termination codon at base 8,372-8,374, and could encode a peptide of 881 amino acids. There is also a termination codon at base 7,934. Based on protein alignments with the HTLV-III_B *env* gene¹⁶, this termination codon is located in the transmembrane portion of the envelope gene immediately before the splice acceptor site in the third exon of *tat3* mRNA. The homology between STL_V-III_{AGM} and HTLV-III_B decreases considerably after the termination codon at position 7,934 of the STL_V-III_{AGM} genome. We have observed a termination codon at 7,934 in several independent STL_V-III_{AGM} clones and in a biologically active³⁰ HTLV-IV clone (ref. 11 and J. Mullins, personal communication). Although such a termination codon is not reported in the published sequence of HIV-2_{ROD}, it appears to be present in some HIV-2_{ROD} clones in the same position (Alizon, M., personal communication). The envelope proteins of STL_V-III_{AGM} and HIV-2_{ROD} are 70% identical in both the extracellular and the transmembrane portion of the protein, with higher homology between the cleavage site and

Table 1 Per cent amino-acid homologies between virus isolates

<i>gag</i>		<i>pol</i>		<i>env</i>		<i>sor</i> (Q)	3' <i>orf</i> (F)	Weighted average
p17	p24	protease	RT	large <i>env</i>	small <i>env</i>			
STLV-III _{AGM} and HIV-2 _{ROD}								
	82		76		70	64	56	72
84	88	88	80	71	70			
STLV-III _{AGM} and HIV-1 (HTLV-III _B)								
	51		53		34	24	34	43
HIV-2 _{ROD} and HIV-1 (HTLV-III) _B								
	52		55		35	28	20	43

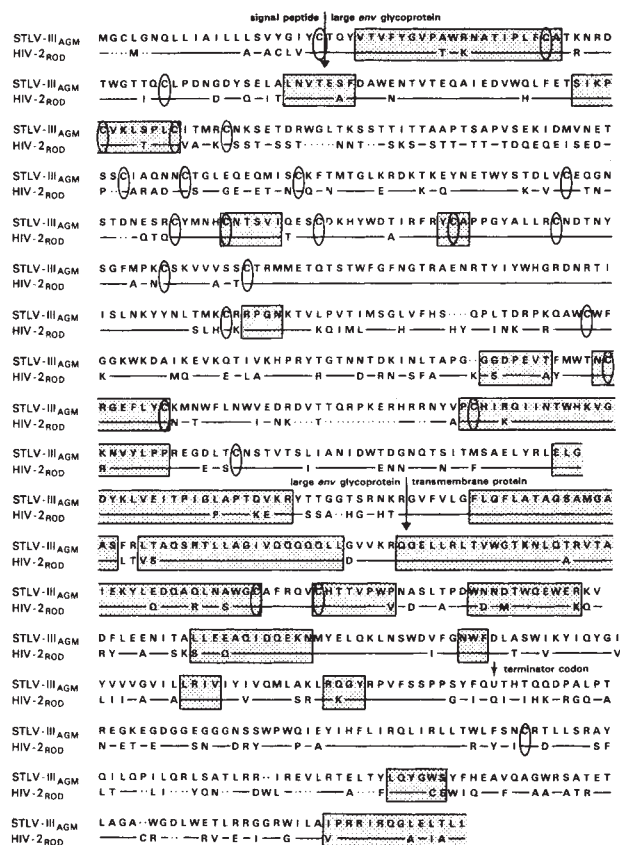


Fig. 3 Similarities between the *env* protein products of HIV-2_{ROD} and STLV-III_{AGM}. The alignment of the envelope proteins was derived using the algorithm program of Dayhoff and colleagues. Arrow, putative cleavage site between the extracellular and transmembrane portions of the precursor envelope protein. Elipses, conserved cysteines. Homology is indicated by lines. Shadowed areas, regions conserved also in the HIV-1 isolates. Dots indicate deletions in HIV-2_{ROD} relative to STLV-III_{AGM}.

the termination codon (82%) than at the carboxy terminus of the envelope open reading frame (54%). Deletion of part of the carboxy terminus of the transmembrane portion of the HTLV-III_B envelope gene reduces the cytopathic effect *in vitro*³¹. The presence of this termination codon in STLV-III_{AGM}, HTLV-IV and HIV-2, which would eliminate 40% of the transmembrane protein, could modulate the pathogenicity of the virus.

An alignment of the inferred amino acid sequences of the STLV-III_{AGM} and HTLV-III_B *env* gene shows 38% homology overall (Fig. 3), and permits us to predict that the signal peptide extends to the serine at bases 5,810–5,812 and that the large and small envelope proteins are cleaved, as in HTLV-III_B³², between the arginine and glycine residues at bases 7,307–7,312. Based on this assumption, the homologies of the two large *env* proteins and transmembrane proteins are 35% and 44% respectively; the hydropathy profiles for the transmembrane proteins are almost identical, consistent with functional similarity. Much higher homology can be found between the envelope proteins of STLV-III_{AGM} and HIV-2_{ROD} (see Table 1).

Of the 22 cysteines present in the HTLV-III_B *env* protein, 21 are found in the same position in the STLV-III_{AGM} *env* protein (not shown). Different strains of HIV-1, which vary by 20–25% in the sequence of the *env* protein, similarly conserve these cysteines, indicating that disulphide bridges are critical. Similarly, 26 of 27 cysteines are conserved between the envelope proteins of STLV-III_{AGM} and HIV-2_{ROD} (Fig. 3). The localized homologies between the *env* genes of STLV-III_{AGM} and HIV-1 and HIV-2_{ROD} occur in those regions which are most conserved among different strains of HIV-1, suggesting that the conserved

regions are functionally important and perhaps involved in binding the CD4 protein. It has recently been shown by Lasky *et al.* (personal communication) that the region in the HIV-1 gp120 that contains the last cysteine is critical for binding to the CD4 molecule (amino acids 442–463 in STLV-III_{AGM}, which are conserved in HTLV-III_B, STLV-III_{AGM} and HIV-2_{ROD}; Fig. 3). Cease *et al.*³³ have found that one of two peptides which elicit T cell immunity maps within the region of the envelope may bind the CD4 molecule.

An open reading frame encoding 96 amino acids can be found (nucleotides 7,900–8,187) following immediately after a splice junction present in a functional STLV-III_{AGM} *tat* cDNA (Colombini-Hatch *et al.*, personal communication). The probable protein encoded by this region shares 29% homology with an equivalent open reading frame in HIV-1, designated *art*³⁴ or *trs*³⁵. The protein product of HIV-1 *trs* appears to be essential for the synthesis of gp120³⁶.

The last large open reading frame (8,196–8,828) has a coding potential for 211 amino acids and a potential initiating AUG codon at position 8,208; the overall amino acid homology between the putative STLV-III_{AGM} protein and the HTLV-III_B 3'*orf* is 34%, with a highly conserved region (around 70% homology) spanning about 80 amino acids. The hydropathy profile of this region in the different viruses is almost indistinguishable. It is likely that these 80 amino acids are the core of the functional domain of the 3' *orf* protein.

Thus the molecular characteristics of STLV-III reveal that its genetic organization is very similar to HIV-1 and HIV-2. STLV-III_{AGM} and HIV-2 are more closely related to each other and more distantly related to HIV-1, but the extent of similarity among these primate viruses indicates that they arose from a common ancestor. The close relationship of STLV-III to the various HIV-1 isolates and even closer relationship of STLV-III_{AGM} to the several recent isolates of West African retroviruses suggests that antibody blood tests for one could detect all members of this group.

We thank Drs P. J. Kanki and M. Essex for providing the STLV-III_{AGM} infected cell line K6W, C. A. Wilson for help with some of the experiments and E. Vinar and M. K. Meyer for editorial assistance. Centocor (Malvern PA) provided the salary for G.F. Also supported in part by the P.F. Oncologia (CNR), the Italian-American Bilateral Agreement for Cancer Research (NCI) and the Yamaguchi Yoshida Cancer Study Grant (UICC). *Note added in proof:* The sequence data have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00295.

Received 2 April; accepted 24 June 1987.

- Barre-Sinoussi, F. *et al.* *Science* **220**, 868–871 (1983).
- Popovic, M. *et al.* *Science* **224**, 497–499 (1984).
- Gottlieb, M. S. *et al.* *N. Engl. J. Med.* **305**, 1425–1429 (1981).
- Daniel, M. D. *et al.* *Science* **228**, 1201–1204 (1985).
- Letvin, N. L. *et al.* *Science* **230**, 71–73 (1985).
- Kanki, P. J., Alroy, J. & Essex, M. *Science* **230**, 951–954 (1985).
- Murphy-Corb, M. *et al.* *Nature* **321**, 435–437 (1986).
- Fultz, P. N., *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 5286–5290 (1986).
- Kanki, P. J., *et al.* *Science* **232**, 238–242 (1986).
- Franchini, G. *et al.* in *AIDS Research and Human Retroviruses* **3**, 11–17 (1987).
- Kornfeld, H., Riedel, N., Viglianti, A., Hirsch, V. & Mullins, J. I. *Nature* **326**, 610–613 (1987).
- Clavel, F. *et al.* *Science* **233**, 343–346 (1986).
- Albert, J. *et al.* in *AIDS Research and Human Retroviruses* **3**, 2–10 (1987).
- Clavel, F. *et al.* *Nature* **324**, 691–695 (1986).
- Guyder, M. *et al.* *Nature* **326**, 662–669 (1987).
- Ratner, L. *et al.* *Nature* **313**, 277–284 (1985).
- Sanchez-Pescador, R. *et al.* *Science* **227**, 484–492 (1985).
- Muesing, M. A. *et al.* *Nature* **313**, 450–458 (1985).
- Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* **40**, 9–17 (1985).
- Ziff, E. B. & Evans, R. M. *Cell* **15**, 1463–1475 (1978).
- Proudfoot, N. J. & Brownlee, G. G. *Nature* **263**, 211–214 (1976).
- Guo, H.-G. *et al.* in *AIDS Research and Human Retroviruses* (in the press).
- Derse, D. *et al.* *J. Virol.* **61**, 743–747 (1987).
- Stephens, R. M., Casey, J. W. & Rice, N. R. *Science* **231**, 589–594 (1986).
- Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3618–3622 (1983).
- Ratner, L. *et al.* in *AIDS Research and Human Retroviruses* (in the press).
- diMarzio Veronese, F. *et al.* *Science* **231**, 1289–1292 (1986).
- Kan, N. C. *et al.* *Science* **231**, 1553–1555 (1986).
- Arya, S. K., Guo, C., Joseph, S. F. & Wong-Staal, F. *Science* **229**, 69–73 (1985).

30. Hirsh, V., Reidel, N. & Mullins, J. *Cell* **49**, 307-319 (1987).
31. Fisher, A. G. *et al. Science* **233**, 655-659 (1986).
32. diMarzo Veronese, F. *et al. Science* **229**, 1402-1405 (1985).
33. Cease, K. B. *et al. Proc. natn. Acad. Sci. U.S.A.* **84**, 4249 (1987).
34. Sodroski, J. *et al. Nature* **321**, 412-417 (1986).
35. Feinberg, M. B., Jarrett, R. F., Aldovini, A., Gallo, R. C. & Wong-Staal, F. *Cell* **46**, 807 (1986).
36. Knight, D. M., Flemerfelt, F. A. & Grayes, J. *Science* **236**, 837-840 (1987).
37. Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. U.S.A.* **74**, 5463-5467 (1977).

Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses

Lisa Chakrabarti*, Mireille Guyader†, Marc Alizon†, Muthiah D. Daniel‡, Ronald C. Desrosiers‡, Pierre Tiollais* & Pierre Sonigo*

* Unité de recombinaison et expression génétique (INSERM U163, CNRS UA271) and † Unité d'oncologie virale (CNRS UA1157), Institut Pasteur, 25 rue du Dr Roux 75724 Paris Cedex 15, France
‡ New England Regional Primate Research Center, Harvard Medical School, Southboro, Massachusetts 01772, USA

Because of the growing incidence of AIDS (acquired immune deficiency syndrome), the need for studies on animal models is urgent. Infection of chimpanzees with the retroviral agent of human AIDS, the human immunodeficiency virus (HIV), will have only limited usefulness because chimpanzees are in short supply and do not develop the disease. Among non-human primates, both type D retroviruses and lentiviruses can be responsible for immune deficiencies. The D-type retroviruses¹⁻³, although important pathogens in macaque monkey colonies, are not satisfactory as a model because they differ in genetic structure and pathophysiological properties from the human AIDS viruses^{4,5}. The simian lentivirus, previously referred to as simian T-cell lymphotropic virus type III (STLV-III), now termed simian immunodeficiency virus (SIV) is related to HIV by the antigenicity of its proteins and in its main biological properties, such as cytopathic effect and tropism for CD4-bearing cells⁶⁻⁹. Most importantly, SIV induces a disease with remarkable similarity to human AIDS in the common rhesus macaques, which therefore constitute the best animal model currently available¹⁰. Natural or experimental infection of other monkeys such as African green monkeys or sooty mangabeys has not yet been associated with disease^{8,9,11}. Molecular approaches of the SIV system will be needed for biological studies and development of vaccines that could be tested in animals. We have cloned and sequenced the complete genome of SIV isolated from a naturally infected macaque that died of AIDS. This SIV_{MAC} appears genetically close to the agent of AIDS in West Africa, HIV-2 (ref. 12), but the divergence of the sequences of SIV and HIV-2 is greater than that previously observed between HIV-1 isolates¹³.

We have previously shown that probes derived from the HIV-2 genome could detect viral DNA in SIV_{MAC} infected cells¹⁴. Cloned subgenomic fragments of HIV-2, representing *gag-pol*, *env* or long terminal repeat (LTR) sequences, were used to screen, in low stringency conditions, a genomic library of HUT-78 cells infected by SIV_{MAC} isolate Mm142-83 (ref. 6). Rhesus monkey Mm142-83 became infected with SIV *in utero* and had constant health problems until her death at ~2 years of age with immunodeficiency and lymphoproliferative syndrome^{6,15}. The nucleotide sequence of one SIV_{MAC}142 clone, λ SIV1, that hybridized with all HIV-2 subgenomic probes, was determined (Fig. 1). The clone λ SIV1 contains one integrated provirus, lacking only the first 257 base pairs (bp) of the left LTR; the right LTR is complete and followed by 7 kilobases (kb) of cellular DNA. Biological activity of this clone (monitored by reverse transcriptase assay) was shown in HUT 78 cells at day 12 after transfection (Y. Naidu, Y. Li, H. Kester and G. Jaenel, unpublished results).

The genome of SIV *mac* is 9,643 nucleotides long (in its RNA form). The organization of its open reading frames, 5' LTR-*gag-pol*-central region-*env*-F-3'LTR (Fig. 2), is typical of a lentivirus¹⁶ and similar in its basic features to that of the HIVs¹². The only remarkable difference in the genetic organizations of SIV, HIV-1 and HIV-2 resides in the open reading frame (ORF) we termed X, absent in HIV-1¹⁷ and differently positioned relative to ORF R in SIV and HIV-2.

Nucleotide sequence comparison of SIV, HIV-1 (ref. 17) and HIV-2 (ref. 12) reveals considerable homology between SIV and HIV-2. These two viruses share about 75% overall nucleotide sequence homology, but both of them are only distantly related to HIV-1 with about 40% overall homology.

The restriction map of the SIV_{MAC}142 isolate, derived from the nucleotide sequence of λ SIV1 is similar to those previously established by molecular cloning and Southern blotting^{18,19} for the remarkably related viruses apparently originating from West Africa and isolated from African green monkeys (STLV-III_{AGM}), or from healthy individuals (human T-cell lymphotropic virus type IV, HTLV-IV). Of 31 sites, 27 are conserved (87%) between λ SIV1 and the STLV-III/HTLV-IV consensus map. This finding is surprising when one considers the variability of HIV-1 isolates and the fact that less than 30% of the sites are conserved between λ SIV1 and the ROD isolate of HIV-2 (ref. 12), from which HTLV-IV is serologically indistinguishable. Other analyses^{19,20} strongly suggest that the original isolates of STLV-III_{AGM} and HTLV-IV are laboratory acquired contaminants with SIV_{MAC}251 which was obtained from the same closed colony as the isolate we studied²⁰. To support this idea, comparison of SIV_{MAC}LTR with published partial LTR sequences of STLV-III_{AGM} or HTLV-IV¹⁹ reveals a homology of over 98%.

The SIV LTRs are 831 bp long, and by alignment to those of

Table 1 % Homology between retroviral proteins

<i>a</i>	<i>Gag</i>	<i>Pol</i>	<i>Env</i>							
SIV _{MAC} compared with			EGP	TMP	<i>F</i>	<i>Q</i>	<i>R</i>	<i>X</i>	<i>Tat</i>	<i>Art</i>
HIV-2 _{ROD}	87	83.6	75.4	74.1	59.8	73.4	70.3	85.7	59.2	61.0
HIV-1 _{BRU}	57.3	59	34.0	49.1	45.7	37.8	57	—	48.1	33.3
HIV-2 _{ROD} compared with										
HIV-1 _{BRU}	57.7	59.4	39.4	44.8	37.7	34.6	52.2	—	42.8	44.8
<i>b</i>	HIV-2 _{ROD}	HIV-1 _{BRU}	HIV-1 _{ELI}	HIV-1 _{MAL}	VISNA	EIAV	MPMV	HTLV-1	RSV	
SIV _{MAC}	83.6	59	58.2	59.2	42.7	44.5	34	34.7		35.7
HIV-2 _{ROD}	—	59.4	61.6	59	43.7	43.8	37.8	34.8		35.9
HIV-1 _{BRU}	59.4	—	94.4	92	41.9	42.2	36.4	33.3		34.5

Alignments were performed using the program NUCALN³¹ with default parameters. The number indicates the % of amino-acid identity in the aligned domains, that is excluding the regions of insertion/deletion. a, Comparisons of all proteins of HIV-1 (ref. 17), HIV-2 (ref. 12) and SIV. b, Comparisons of *pol*-encoded proteins of different retroviruses. HIV-1_{ELI} and HIV-1_{MAL} are Zairian isolates of HIV-1 (ref. 13). EIAV: equine infectious anaemia virus³². MPMV: Mason Pfizer monkey virus⁴. HTLV-1: Human T-cell leukaemia virus type I (ref. 33). RSV: Rous sarcoma virus³⁴.

Fig. 1 Left and opposite, the complete nucleotide sequence of the SIV_{MAC} genome. The 9,643 nucleotides are numbered from the cap site to the polyadenylation site and shown together with the deduced amino-acid sequences. PBS, primer binding site complementary to lysine tRNA. Ppt1 and ppt2 are the two copies of the polypurine tract. In *U3* are shown the two enhancer-like sequences (e, underlined, ref. 35), the Sp1-like binding sites (Sp1, overlined) and the promoter (TATAAA, boxed). The genes are translated from the first AUG of the ORF, except for *pol* where the whole ORF is translated. Filled circles, termination codons. SD and SA indicate the splice donor and acceptor of the intron of the *tat* and *art* genes. The in-frame stop codon found in the *env* gene at position 8,298 is indicated by three asterisks. **Methods.** Molecular cloning of SIV proviral DNA was carried out with Sau3A partial digest of DNA from HUT78 cells infected with isolate 142-83 (ref. 6) and λ EMBL3 arms (supplied by Vector Cloning Systems). Plaques were screened *in situ* using nick-translated subgenomic fragments of HIV-2 (ref. 14) as probes. Stringency of hybridization, 50 °C in 5 × SSC and washing 50 °C in 2 × SSC. Sequencing of proviral DNA from λ SIV1 was carried out using standard M13 shotgun and dideoxy nucleotides procedures as already described¹⁶.

[illegible]

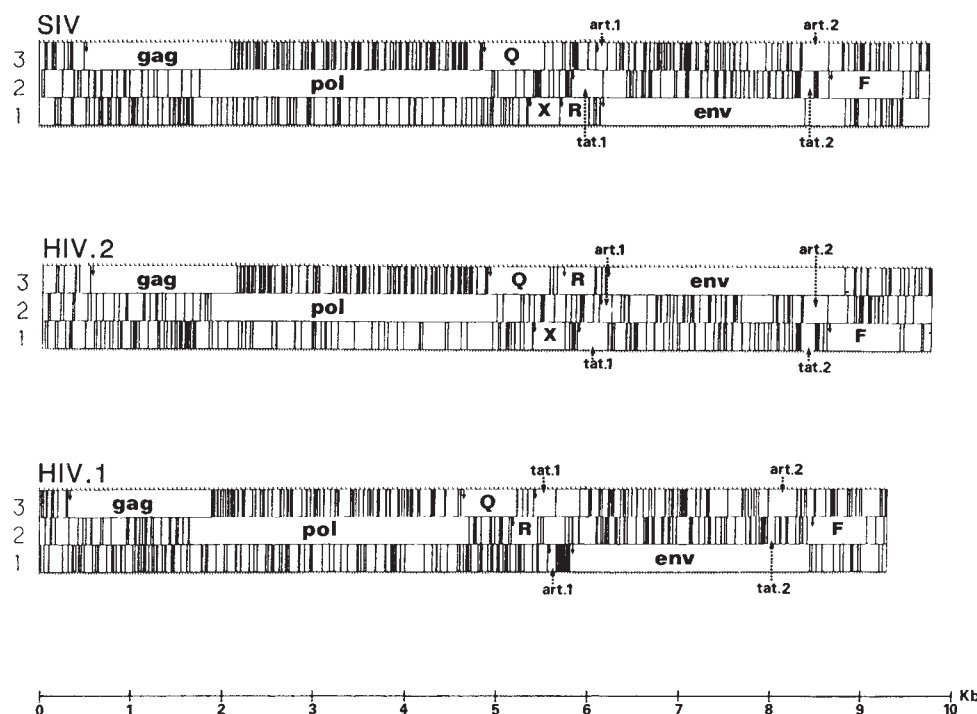


Fig. 2 Genomic organization of SIV compared to HIV-1 (BRU isolate, ref. 17) and HIV-2 (ROD isolate, ref. 12). Vertical bars represent the stop codons in the three reading frames. Arrows indicate the potential AUG initiator codon of each open reading frame. Tat1&2 and art1&2 designate the open reading frames containing the first and second coding exons of the *tat* and *art* genes.



Fig. 3 The alignment of the SIV_{MAC}, HIV-2_{ROD} (ref. 12) and HIV-1_{BRU} (ref. 17) envelope proteins. Gaps were introduced to optimize the alignment. Asterisks indicate amino acid identity. Potential cleavage sites are boxed and shown by arrows. Cysteines are boxed, potential N-glycosylation sites are overlined. The filled square in the SIV_{MAC} sequence corresponds to an in-frame nonsense codon. EGP, external glycoprotein; TMP, transmembrane protein.

HIV-2, the length of the internal domains of the SIV LTR have been estimated as: *U3*, 514; *R*, 175; *U5*, 142. The nucleic-acid homology with HIV-2 LTR is 66% in *U3*, 70% in *U5* and 95% in *R*. This important level of LTR homology is associated with conservation of transcription signals and secondary structures common to HIV-1 and HIV-2 LTRs¹².

The *gag* precursor of SIVmac has a calculated relative molecular mass (M_r) of 58.9 K, comparable with the mass of the p55 antigen, estimated from SDS gels; this precursor is likely to be processed into the proteins designated p16 (amino terminus, calculated M_r = 15.3 K), p27 (major core protein) and p12 (carboxy-terminal, nucleic-acid binding protein^{11,21,22}). The N-terminal extremity of the p27⁸⁹⁸ of SIV isolated from a macaque (*Macaca nemestrina*) at the Washington Primate Research Center (MnIV/WRPC) was sequenced previously²³ and only 1 difference out of 23 amino acids is observed with SIV_{MAC142} (Val5 > Ile). In the same region, 2 differences have been found with HIV-2 and 10 (with 1 insertion) with HIV-1.

The SIV *pol* ORF (*pol* protein product, total M_r = 119.6 K) probably encodes the p64 and p53 antigens (reverse transcriptase and endonuclease²²). The SIV *env* ORF could encode the gp160, gp120 and gp32 antigens^{11,21,22}: the precursor (p env), the external glycoprotein (EGP) and the transmembrane protein (TMP). The calculated M_r of non-glycosylated products are p env, 101 K; EGP, 60 K; TMP, 40.7 K; the numbers of potential N-glycosylation sites are: EGP, 22; TMP, 4 (Fig. 3).

An in-frame nonsense codon interrupts the *env* ORF at nucleotide 8,298 in clone λ SIV1. Strikingly, the same stop codon (TAG) was found at exactly the same position in the envelope of an integrated genome of HIV-2, but not (replaced by CAG) in complementary DNA derived *in vitro* from viral RNA of the same isolate¹². Such stop codons experimentally introduced after the hydrophobic part of the TMP (in the so-called cytoplasmic tail) in Rous sarcoma virus do not impair retroviral replication *in vitro*²⁴. In SIV, as in HIV-2, the stop codon marking the beginning of the *env* ORF matches with the splice donor site of the first coding exons of the *tat* and *art* genes. Also, the in-frame stop matches with the splice acceptor site for the second coding exons of these genes. The part of the *env* ORF that corresponds to the second intron of the *tat* and *art* regulating genes is thus exactly delimited by two stop codons. The existence of these stop signals could influence (or result from) the differential splice that gives rise alternatively to *tat* and *art* (in a regulating-like phase) or *env* (in a structural-like phase) and may modulate the viral gene expression and pathogenicity, as recently observed for visna virus (ref. 25 and R. Vigne, personal communication). Such codons may also account for the variations in the reported size^{22,26,27} of the TMP for HIV-2 or SIV: gp40, gp36 or gp32 (calculated M_r : total TMP, 41K; from cleavage site to stop, 24K).

Comparisons of the proteins of SIV with those of HIV-1 and 2, summarized in Table 1a, quantify the relatedness of these viruses. Although SIV and HIV-2 appear closely related, their observed level of divergence (~15% in *gag* and *pol* and ~30% in *env*) remains higher than that observed for the most distant isolates of HIV-1 (maximum ~10% in *pol* and ~20% in *env*¹³). Thus, classification of HIV-1, HIV-2 and SIV into either two or three subgroups of primate lentiviruses should remain an open issue at this time. Continued sequence analysis with additional authentic isolates will be needed to determine the extent to which the SIV and HIV-2 groups overlap and the genetic variability of this group of viruses. In any case, it already seems that the conserved regions defined by comparison of sequenced isolates of HIV-1 and HIV-2 (refs 12, 13) are also conserved in SIV (Fig. 3).

As already observed for different isolates of HIV-1 and HIV-2 (refs 12, 13), the transacting *tat* genes of HIV-2 and SIV show a large degree of divergence (40.8%). But preliminary experiments have shown that the clone λ SIV1 encodes a functional transactivator with a similar specificity to that of HIV-2 *tat*—that

is, it transactivates the HIV-2 LTR more efficiently than the HIV-1 LTR (ref. 12 and M. Emerman *et al.*, manuscript in preparation).

Theories regarding the possible existence of an animal reservoir for human AIDS and its role in the recent apparition of the AIDS pandemics must now take into account the sequence relationships of the lentiviruses and the fact that at least two divergent groups of lentiviruses are found in the human population. The observation that the divergence of all known lentiviruses (Table 1b) approximately follows the divergence of the infected species does not support a recent horizontal inter-species transmission. Sequence comparisons (Table 1b) reveal also that HIV-2 or African isolates of HIV-1, although obviously derived from a common ancestor, do not appear simply as evolutionary intermediates between SIV and European or US isolates of HIV-1. Thus, even if it is supposed that monkeys could be an accessory reservoir for HIV-2, available sequence data are inconsistent with the idea that AIDS emerged from recent transmission of SIV to humans, followed by rapid viral evolution toward HIV-1.

Recombinant DNA techniques have allowed efficient production of a variety of HIV-1 antigens (see for example refs 28–30) but adequate testing of such antigens as potential vaccines is limited by the absence of systems for live virus challenge. The availability of sequenced molecular clones should now permit the synthesis of similar products of SIV origin. These SIV antigens can be tested reliably in macaques, before testing analogous HIV antigens in humans.

We thank Drs F. Clavel, M. Emerman, L. Montagnier and S. Wain-Hobson for helpful discussions. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00277.

Note added in proof: Following the submission of this paper, Hirsch *et al.* have reported a partial sequence of STL-3_{AGM} (Cell 49, 307–319; 1987). Their findings are similar to ours, in particular regarding the existence of an in-frame stop codon in the *env* gene at the same position as in our sequence. By comparison of the envelope proteins, STL-3_{AGM} appears to be closely related to SIV_{MAC} (91.4% of amino-acid identity).

Received 18 May; accepted 4 June 1987.

1. Daniel, M. D. *et al.* *Science* **223**, 602–605 (1984).
2. Marx, P. A. *et al.* *Science* **223**, 1083–1086 (1984).
3. Stromberg, K. *et al.* *Science* **224**, 289–292 (1984).
4. Sonigo, P., Barker, C., Hunter, E. & Wain-Hobson, S. *Cell* **45**, 375–385 (1986).
5. Power, M. D. *et al.* *Science* **231**, 1567 (1986).
6. Daniel, M. D. *et al.* *Science* **228**, 1201–1204 (1985).
7. Kanki, P. J., Alroy, J. & Essex, M. *Science* **230**, 951–954 (1985).
8. Fultz, P. N. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 5286–5290 (1986).
9. Murphy-Corb, M. *et al.* *Nature* **321**, 435–437 (1986).
10. Letvin, N. L. *et al.* *Science* **230**, 71–73 (1985).
11. Kanki, P. S. *et al.* *Lancet* **i**, 1330–1332 (1985).
12. Guyader, M. *et al.* *Nature* **326**, 662–669 (1987).
13. Alizon, M., Wain-Hobson, S., Montagnier, L. & Sonigo, P. *Cell* **46**, 63–74 (1986).
14. Clavel, F. *et al.* *Nature* **324**, 691–695 (1986).
15. Chalifoux, L. V. *et al.* *Lab. Invest.* **55**, 43–50 (1986).
16. Sonigo, P. *et al.* *Cell* **42**, 369–382 (1985).
17. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. *Cell* **40**, 9–17 (1985).
18. Hirsch, V. *et al.* *Proc. natn. Acad. Sci. U.S.A.* **83**, 9754–9758 (1986).
19. Kornfeld, H., Riedel, N., Viglianti, G. A., Hirsch, V. & Mullins, J. I. *Nature* **326**, 610–613 (1987).
20. Desrosiers, R. C., Daniel, M. D., Letvin, N. L., King, N. W. & Hunt, R. D. *Nature* **327**, 107 (1987).
21. Kanki, P. J. *et al.* *Science* **228**, 1199–1201 (1985).
22. Kanki, P. J. *et al.* *Science* **232**, 238–243 (1986).
23. Benveniste, R. F. *et al.* *J. Virol.* **60**, 483–490 (1986).
24. Perez, L. G., Davis, G. L. & Hunter, E. *J. Virol.* (in the press).
25. Davis, J. L., Molineaux, S. & Clements, J. E. *J. Virol.* **61**, 1325–1331 (1987).
26. Brun-Vezinet, F. *et al.* *Lancet* **ii**, 128–132 (1987).
27. Clavel, F. *et al.* *Science* **233**, 343–346 (1986).
28. Zarling, J. M. *et al.* *Nature* **323**, 344–346 (1986).
29. Putney, S. D. *et al.* *Science* **234**, 1392–1395 (1986).
30. Lasky, L. *et al.* *Science* **233**, 209–212 (1986).
31. Wilbur, W. J. & Lipman, D. J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 726–730 (1983).
32. Kawakami, T. *et al.* *Virology* **156**, 853–869 (1987).
33. Seiki, M., Hattori, S., Hiriyama, Y. & Yoshida, M. *Proc. natn. Acad. Sci. U.S.A.* **80**, 3618–3622 (1983).
34. Schwartz, D. E., Tizard, R. & Gilbert, W. *Cell* **32**, 853–869 (1983).
35. Nabel, G. & Baltimore, D. *Nature* **326**, 711–713 (1987).

New human and simian HIV-related retroviruses possess functional transactivator (*tat*) gene

Suresh K. Arya*, Barbara Beaver*, Linda Jagodzinski†, Barbara Ensoli*, Phyllis J. Kanki‡, Jan Albert§||, Eva-Maria Fenyo||, Gunnel Biberfeld§, Jean F. Zagury¶, Francoise Laure¶, Myron Essex†, Erling Norrby||, Flossie Wong-Staal* & Robert C. Gallo*

* Laboratory of Tumor Cell Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA

† Biotech Research Laboratories, Rockville, Maryland 20850, USA

‡ Department of Cancer Biology, Harvard School of Public Health, Boston, Massachusetts 02115, USA

§ The National Bacteriological Laboratory, Stockholm, Sweden

|| Department of Virology, Karolinska Institute, School of Medicine, Stockholm, Sweden

¶ Laboratoire de Biologie Cellulaire, University Pierre et Marie Curie, Paris, France

New human retroviruses antigenically related to HIV and even more closely to STLV-III have been recently isolated from individuals from some West African countries¹⁻⁴. One of these viruses, HTLV-IV_p, was reportedly isolated from lymphocytes of a healthy female prostitute^{1,2}. Another isolate, LAV-2_{FG}, was obtained from an AIDS patient and third, SBL-6669, from an individual with lymphadenopathy⁴. Current epidemiological studies indicate that some of these virus isolates cause immune deficiency whereas others may not or may be less efficient at inducing immune deficiency. Similarly, STLV-III apparently does not cause immune deficiency in its natural host, African green monkey⁵. A novel feature of HIV is the possession of a gene termed *tat*, which is implicated in its pathobiology⁶⁻¹². We report here that, like HIV, HTLV-IV_p, LAV-2_{FG} (HIV-2) and SBL-6669, as well as STLV-III_{AGM} possess the putative *tat* gene, irrespective of their pathogenic potential *in vivo*. Interestingly, HTLV-IV_p/LAV-2_{FG} long terminal repeat (LTR) is equally well transactivated by the HTLV-IV_p/LAV-2_{FG} and HTLV-III_B *tat* function, HTLV-III_B LTR responds better to its own *tat* function.

The human T-lymphotropic retroviruses (HTLVs) share many properties, including preferred tropism for T lymphocytes bearing the CD4 (T4) antigen, similar modes of transmission and, probably, African origin. They also differ significantly from each other, forming two distinct groups. HTLV-I¹³⁻¹⁶, a transforming virus, is the prototype of the first group. It causes adult T-cell leukaemia (ATL) and appears to be involved in B-cell malignancies and in central nervous system disease^{17,18}. This group also includes HTLV-II¹⁹ and variants of HTLV-I²⁰ which are associ-

ated with T-cell malignancies different from ATL. Human immune deficiency virus (HIV)²¹⁻²⁴, also known as HTLV-III or lymphadenopathy virus (LAV), a non-transforming virus, is the prototype of the second group. It causes acquired immune deficiency syndrome (AIDS) and nervous system disorders. Recently, new members of this group of non-transforming human retroviruses were isolated from individuals from different countries of West Africa¹⁻⁴. These isolates show major antigenic characteristics that distinguish them from HIV, although they retain many features in common¹⁻⁴. Remarkably, counterparts of both groups of human retroviruses also exist in old-world subhuman primates. They have been termed simian T-lymphotropic virus-I and -III (STLV-I and -III). STLV-III, related to HIV and even more closely to HTLV-IV¹⁻⁵ does not appear to cause AIDS in its natural host, the African green monkey or related species.

To test if the new non-transforming human retroviruses contained a transactivating gene, we first cloned HTLV-IV_p, LAV-2_{FG} and STLV-III_{AGM} LTRs by way of cDNA cloning. Several independent clones containing 3'-LTR and sequences extending upstream were characterized and sequenced. After this manuscript was submitted, the sequences of the LTRs of HIV-2²⁵, HTLV-IV and STLV-III²⁶ and the complete sequence of the HIV-2 genome²⁷ were published. The sequences of our cDNA clones of these viruses are very similar or identical to the published reports. Differences of a few nucleotides, where observed, probably reflect strain differences. As with Kornfeld *et al.*²⁵, we find the sequences of HTLV-IV and STLV-III to be more than 95% homologous. This may reflect inter-species transmission of the same or nearly identical virus(es). Cross-species transmission of retroviruses can occur under certain circumstances. For example, STLV-III infection of macaque monkeys may be a result of inter-species transmission from infected African green or mangabey monkeys kept in close proximity²⁶⁻³⁰. Similarly, HIV can be transmitted to chimpanzees³¹ and HTLV-I to certain monkeys³². Kornfeld *et al.*²⁵ have also noted that HTLV-IV and STLV-III may not be independent virus isolates but result from transmission of the same virus to different cell cultures maintained in the laboratory. This is possible but it is equally possible that the transmission occurred in nature. Future studies with more human and simian virus isolates may clarify this issue.

For transactivation experiments, cloned HTLV-IV_p DNAs (V17 and V8) and LAV-2_{FG} DNA (LV-2) were placed upstream of the bacterial chloramphenicol acetyltransferase (CAT) gene carried on a vector lacking promoter-enhancer sequences (pSVOCat)²⁸ and the resulting plasmids were termed pV17Cat, pV8Cat and pLV2Cat. Two additional plasmids, pSV2Cat³³ and pC15Cat^{8,9} respectively contained SV40 promoter-enhancer and HTLV-III 3'-LTR linked to the CAT gene. The plasmid DNAs were transfected into virus-infected and uninfected human HUT 78 cells by the DEAE-dextran protocol and transactivation of the LTR-linked CAT gene was measured by conversion of

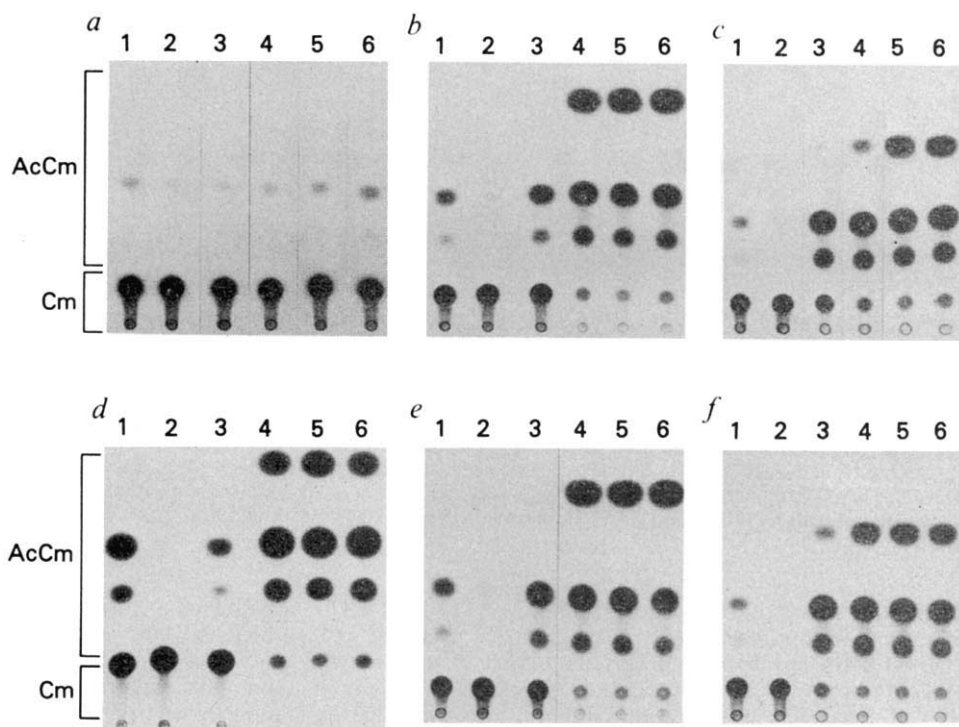
Table 1 Transactivation of HTLV-IV, LAV-2 and HTLV-III LTRs in virus-infected cells

Plasmid	Regulatory elements	% Conversion of chloramphenicol to acetylated forms					
		HUT78*	HTLV-IV _p /HT*	SBLV6669/HT*†	LAV-2 _{FG} /HT*	STLV-III _{AGM} /HT*	HTLV-III _B /H9*
pSV2Cat	SV40	2.4 ± 1.7	3.0 ± 0.35	0.7 ± 0.6	9.5 ± 1.0	4.5 ± 3.8	1.3 ± 0.3
pSVOCat	None	0.12 ± 0.01	0.12 ± 0.01	0.14 ± 0.03	0.16 ± 0.02	0.15 ± 0.03	0.14 ± 0.01
pC15Cat	HTLV-III _F	0.43 ± 0.3	11.4 ± 1.9	21.5 ± 34.2	4.6	16.0 ± 9.1	96 ± 1.4
pV17Cat	HTLV-IV _p	0.86 ± 0.4	97.4 ± 2.4	56.5 ± 58.0	98.1 ± 0.2	99.2	99 ± 0.1
pV8Cat	HTLV _p	1.3 ± 0.8	98.7 ± 0.2	56.7 ± 39.0	98.3	98.0 ± 2.0	98 ± 0.8
pLV2Cat	LAV-2 _{FG}	—	99.1	—	96.9	—	99.2
pV17Cat(nc)	HTLV-IV _p	0.13 ± 0.8	0.2 ± 0.05	—	—	—	—

* Mean and standard deviation for 2 to 4 transfection experiments, where shown.

† Relatively wide variation and/or lower values are probably related to the recent infection of the host cells by this virus. The general metabolism and viability of the culture and the proportion of infected cells in the population may fluctuate during early periods.

Fig. 1 Activation of HTLV-IV_P and HTLV-III_B LTR-linked *cat* gene in *a*, uninfected HUT78 cells; *b*, cells infected with HTLV-IV_P; *c*, with SBLV-6669; *d*, with LAV-2_{FG}; *e*, with STLTV-III_{AGM} and *f*, with HTLV-III_B. Lanes 1-6 are for pSV2Cat(SV40), pSVOCat(none), pC15Cat0HTLV-III_A, pV17Cat(HTLV-IV_P), pV17bCat(HTLV-IV_P) and pV8Cat(HTLV-IV_P), respectively. About 10×10^6 cells were transfected by the DEAE-dextran protocol and assays were performed as described before^{11,12}. Cm, chloramphenicol; AcCm, acetylated chloramphenicol.



chloramphenicol to its acetylated forms catalysed by the enzyme in cellular extracts 44-48 h after transfection as previously described^{8,9}. Representative results are shown in Fig. 1 and more extensive data are presented in Table 1.

The HTLV-IV_P and LAV-2_{FG} LTRs contain promoter-enhancer sequence elements, as expected. The CAT gene activity in uninfected HUT78 cells transfected with plasmid DNA with HTLV-IV_P LTR linked in correct orientation to the CAT gene (for example, pV17Cat) was higher than that for the incorrect orientation (for example, pV17Cat-nc) or for the control plasmid lacking promoter-enhancer (pSVOCat; Fig. 1). The CAT gene activity under the control of HTLV-IV_P LTR in HUT 78 cells infected with HTLV-IV_P, LAV-2_{FG} and SBL-6669 and with STLTV-III_{AGM} and HTLV-III_B, was markedly elevated compared to uninfected cells (Fig. 1, Table 1). Similar results were obtained for LAV-2_{FG} LTR. This indicates that the virus-infected cells contain a factor(s) which acts *in trans*, amplifying the expression

of the Cat gene linked to HTLV-IV_P and LAV-2_{FG} LTRs. HTLV-IV_P and LAV-2_{FG} LTRs responded to the transactivating factor in cells infected with all three of the HTLV-IV_P related viruses as well as those infected with STLTV-III_{AGM} and HTLV-III_B. These results suggest that (1) all new isolates (HTLV-IV_P, LAV-2_{FG} and SBL-6669) as well as STLTV-III_{AGM} contain a transactivating gene, (2) their putative transactivating genes are structurally and/or functionally homologous, (3) their LTRs are also structurally and/or functionally related and (4) the HTLV-IV_P putative transactivating gene is related to the similar gene in HTLV-III_B. The fact that HTLV-IV_P and LAV-2_{FG} LTRs are transactivated equally well in cells infected with HTLV-IV_P, LAV-2_{FG}, SBL-6669, STLTV-III_{AGM} and HTLV-III_B whereas HTLV-III_B LTR is more highly transactivated in cells infected with HTLV-III_B than other viruses indicates that HTLV-IV_P/LAV-2_{FG}/SBL-6669/STLTV-III_{AGM} transactivating genes and LTRs are functionally more similar to each other than to

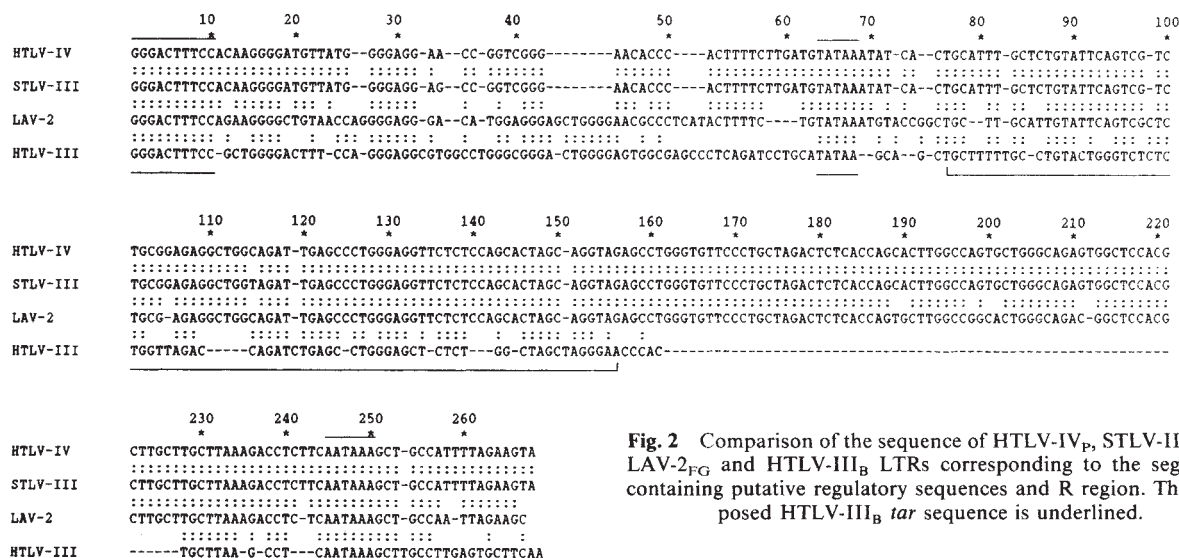


Fig. 2 Comparison of the sequence of HTLV-IV_P, STLTV-III_{AGM}, LAV-2_{FG} and HTLV-III_B LTRs corresponding to the segments containing putative regulatory sequences and R region. The proposed HTLV-III_B *tar* sequence is underlined.

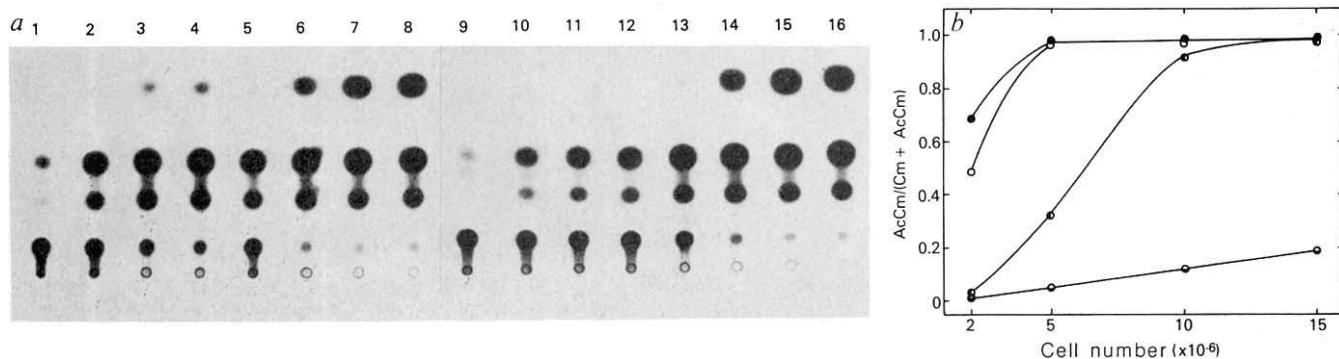


Fig. 3 Comparative activation of HTLV-IV_p and HTLV-III_B LTRs by HTLV-IV_p and HTLV-III_B transactivating functions. Indicated number of virus infected cells were transfected with 10 µg of LTR-Cat plasmid DNAs and the cells were processed to obtain 100 µl of cytoplasmic extracts. Aliquots of 20 µl were used for cat assays^{11,12} and incubation was for 60 min. Lanes 1-4, HTLV-III_B-infected cells transfected with pC15Cat; lanes 5-8, HTLV-III_B-infected cells transfected with pV17Cat; lane 9-12, HTLV-IV_p-infected cells transfected with pC15Cat; lanes 13-16, HTLV-IV_p infected cells transfected with pV17Cat. Δ , HTLV-IV_p-infected cells transfected with pV17Cat; $\circ$, HTLV-III_B-infected cells transfected with pV17Cat; $\circ$, HTLV-III_B-infected cells transfected with pC15Cat; $\circ$, HTLV-IV_p-infected cells transfected with pC15Cat.

HTLV-III_B. A specific sequence (TAR) in HTLV-III_B LTR responsible for optimal transactivation has been identified and localized between 6 base pairs (bp) and 76 bp downstream of the TATA box³⁴⁻³⁶. Such a specific sequence in HTLV-IV_p and STLV-III_{AGM} LTRs must lie within 62 bp upstream and 202 bp downstream of the TATA box. This is the only sequence contained in the fully active clone V8. Furthermore, the LTR sequences of HTLV-IV_p/STLV-III_{AGM}/LAV-2_{FG} and HTLV-III_B can be aligned such that they are about 65% homologous downstream of the TATA box in the proposed *tar* region of HTLV-III_B (Fig. 2). The overall homology in the regulatory region of the LTRs of the new virus isolates is about 85-95%. Notably, the putative enhancer element (GGGACTTCC) is perfectly conserved and the potential SP1 binding sites are also highly conserved in HIV and HTLV-IV_p/STLV-III_{AGM}/LAV-2_{FG}. In addition, the sequences downstream of the TATA box in HTLV-IV_p/STLV-III_{AGM}/LAV-2_{FG} and HTLV-III_B contain imperfect direct and indirect repeats which impart analogous structures to these LTRs and to the transcripts originating therein.

Interestingly, HTLV-IV_p and LAV-2_{FG} LTRs are equally well transactivated in cells infected with HTLV-IV_p, LAV-2_{FG} and HTLV-III_B, whereas HTLV-III_B is more transactivated in cells infected with HTLV-III_B than with HTLV-IV_p or LAV-2_{FG} (Table 1). The differential response of HTLV-III_B LTR in HTLV-III_B- and HTLV-IV_p or LAV-2_{FG}-infected cells is unlikely to be related to lack of *tat* gene function in HTLV-IV_p- or LAV-2_{FG}-infected cells. HTLV-IV_p and LAV-2_{FG} LTRs are highly transactivated in HTLV-IV_p- and LAV-2_{FG}-infected cells. Furthermore, when the source of the *tat* gene function is varied by using different numbers of cells for transfection, the differential response of HTLV-III_B LTR is maintained (Fig. 3). It may be that the HTLV-IV_p/LAV-2_{FG} LTR transactivator interaction has a broader specificity than HTLV-III_B LTR transactivator interaction. Additionally, promoter-enhancer elements of HTLV-IV_p/LAV-2_{FG} may be stronger than such elements in HTLV-III_B, allowing optimal transactivation even with heterologous HTLV-III_B transactivator.

Some results comparing transactivation of the HIV-2 (LAV-2) and HIV-1 LTRs by the homologous and heterologous *tat* functions were recently reported by Guyader *et al.*²⁷. They also found that whereas HIV-2 LTR responded highly to the *tat* function of HIV-1 as well as to its own *tat* function, HIV-1 responded better to its own *tat* function.

The cells infected with all four virus isolates of the West African subgroup (HTLV-IV, LAV-2, SBL-6669 and STLV-III) contain a factor that activates their LTRs in *trans*, indicating that these viruses possess a functional transactivating gene. The transactivating gene of HIV has been implicated in its

pathobiology. As HTLV-IV_p and STLV-III_{AGM} may not always be pathogenic *in vivo*, either their transactivating genes do not function *in vivo*, an unlikely possibility, or this gene function is not sufficient for pathogenicity. Host factors or other viral factors directly or indirectly may modulate the potential pathogenic role of the transactivating genes of these viruses. The interaction between HTLV-IV/STLV-III/LAV-2 (HIV-2) LTRs and the *tat* gene product apparently has broader specificity than the HTLV-III (HIV) LTR: *tat* interaction.

We thank M. M. Tsai and V. Braddon for assistance with DNA sequencing, Drs B. Hahn and G. M. Shaw and colleagues for confirmatory sequencing, Dr S. F. Josephs for useful discussions and Drs P. Marcellin and J. P. Benhamou of Hospital Beaujon, Paris for cooperation. B.E. was supported, in part, by the Italian Association of Cancer Research. P.K. is a Fellow of the Leukemia Society of America. This work was also supported in part by the NIH.

Received 3 April; accepted 1 June 1987.

- Barin, F. *et al. Lancet* ii, 1387-1390 (1985).
- Kanki, P. J. *et al. Science* **232**, 238-243 (1986).
- Clowel, F. *et al. Science* **233**, 343-346 (1986).
- Albert, J. *et al. J. AIDS Res. Human Retroviruses* (in the press).
- Kanki, P. J., Alroy, J. & Essex, M. *Science* **230**, 951-954 (1985).
- Arya, S. K. *et al. Science* **225**, 927-930 (1984).
- Sodroski, J. G. *et al. Science* **227**, 171-173 (1985).
- Arya, S. K., Guo, C., Josephs, S. F. & Wong-Staal, F. *Science* **229**, 69-73 (1985).
- Arya, S. K. & Gallo, R. C. *Proc. natn. Acad. Sci. U.S.A.* **83**, 2209-2213 (1986).
- Sodroski, J. G., Patarca, R., Rosen, C., Wong-Staal, F. & Haseltine, W. A. *Science* **229**, 74-77 (1985).
- Dayton, A. I., Sodroski, J. G., Rosen, C. A., Goh, W. C. & Haseltine, W. A. *Cell* **44**, 941-947 (1986).
- Fisher, A. G. *et al. Nature* **320**, 367-371 (1986).
- Poiesz, B. J. *et al. Proc. natn. Acad. Sci. U.S.A.* **77**, 7415-7419 (1980).
- Miyoshi, I. *et al. Nature* **294**, 770-771 (1981).
- Popovic, M. *et al. Nature* **300**, 63-66 (1982).
- Yoshida, M., Miyoshi, I. & Hinuma, Y. *Proc. natn. Acad. Sci. U.S.A.* **79**, 2031-2035 (1982).
- Wong-Staal, F. & Gallo, R. C. *Nature* **317**, 395-403 (1985).
- Gallo, R. C. in *Cecil's Textbook of Medicine* (eds Smith, L. H. & Wyngaarden, J. B.). 18th edn (Saunders, Philadelphia, in the press).
- Kalyanaraman, V. S. *et al. Science* **218**, 571-573 (1982).
- Hahn, B. H. *et al. Int. J. Cancer* **34**, 613-618 (1984).
- Barre-Sinoussi, F. *et al. Science* **220**, 868-871 (1983).
- Popovic, N., Sarngadharan, M. G., Read, E. & Gallo, R. C. *Science* **224**, 497-500 (1984).
- Levy, J. A. *et al. Science* **225**, 840-842 (1984).
- Salahuddin, S. Z. *et al. Proc. natn. Acad. Sci. U.S.A.* **82**, 5530-5534 (1985).
- Clavel, F. *et al. Nature* **324**, 691-695 (1986).
- Kornfeld, H., Riedel, N., Viglianti, G. A., Hirsch, V. & Mullins, J. I. *Nature* **326**, 610-613 (1987).
- Guyader, M. *et al. Nature* **326**, 662-669 (1987).
- Murphy-Corb, M. *et al. Nature* **321**, 435-437 (1986).
- Kanki, P. J. *et al. Science* **228**, 1199-1201 (1985).
- Daniel, M. D. *et al. Science* **228**, 1201-1204 (1985).
- Alter, J. J. *et al. Science* **226**, 549-552 (1984).
- Yamamoto, N. *et al. Med. microbiol. Immun.* **173**, 57-64 (1984).
- Gorman, C. M., Moffat, L. F. & Howard, B. H. *Molec. cell. Biol.* **2**, 1044-1051 (1982).
- Rosen, C. A., Sodroski, J. G. & Haseltine, W. A. *Cell* **41**, 813-823 (1985).
- Jones, K. A., Kadonaga, J. T., Luciw, P. A. & Tjian, R. *Science* **232**, 755-758 (1986).
- Muesing, M. A., Smith, D. H. & Capon, D. J. *Cell* **48**, 691-701 (1987).

Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin

Shashidhar N. Rao, U. Chandra Singh*, Paul A. Bash* & Peter A. Kollman

Department of Pharmaceutical Chemistry, University of California, San Francisco, California 94143, USA

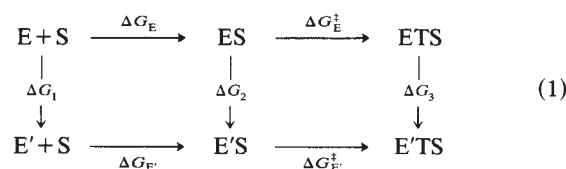
Site-directed mutagenesis is a very powerful approach to altering the biological functions of proteins, the structural stability of proteins and the interactions of proteins with other molecules. Several experimental studies in recent years have been directed at estimating the changes in catalytic properties, (rates of binding and catalysis) in site-directed mutants of enzymes compared to the native enzymes¹⁻³. Simulation approaches to the study of complex molecules have also become more powerful, in no small measure owing to the increase in computer power^{4,5}. These simulations have often allowed results of experiments to be rationalized and understood mechanistically. A new approach called the free-energy perturbation method, which uses statistical mechanics and molecular dynamics can often be used for quantitative calculation of free energy differences⁶. We have applied such a technique to calculate the differential free energy of binding and free energy of activation for catalysis of a tripeptide substrate by native subtilisin and a subtilisin mutant (Asn 155 → Ala 155). Our studies lead to a calculated difference in free energy of binding which is relatively small, but a calculated change in free energy of catalysis which is substantial. These energies are very close to those determined experimentally (J. A. Wells and D. A. Estell, personal communication), which were not known to us until the simulations were completed. This demonstrates the predictive power and utility of theoretical simulation methods in studies of the effects of site-specific mutagenesis on both enzyme binding and catalysis.

In recent years subtilisin has been the subject of crystallographic (refs 7 and 8, and R. Bott and A. Kossiakoff, personal communication) and site-specific mutagenesis experiments¹⁰⁻¹⁴. Site-specific mutagenesis of subtilisin is a particularly appropriate system for theoretical studies because the X-ray structure at high resolution is known and so much binding and kinetic experimental data exist (refs 7-14, and R. Bott and A. Kossiakoff, personal communication). In addition, crystallographic experiments on mutants clearly show that the structural perturbations of mutation are relatively localized (ref. 15 and R. Bott and A. Kossiakoff, personal communication). In the light of the extensive data on subtilisin, we have carried out theoretical simulations using the free energy/molecular dynamics approach.

We chose for our computer simulations the site-specific mutation Asn 155 → Ala 155 for the following reasons. In the crystal structures of a number of serine proteases complexed either with substrate analogues or inhibitors¹⁶⁻¹⁸, the carbonyl oxygen of the reacting amide or ester is stabilized by hydrogen bonding interactions with the polypeptide backbone N-H groups. In trypsin and chymotrypsin, these come from Ser 195 and Gly 193. It has been proposed that these interactions are dominant in stabilizing the transition state enzyme-substrate complexes in reaction pathways of serine proteases. The location of the anionic oxygen of the substrate is called the 'oxyanion hole'. However, in subtilisin, one of the two N-H groups belongs to the side chain of Asn 155 instead of the backbone of a residue (ref. 8, and R. Bott and A. Kossiakoff, personal communication),

whereas the other one is from the backbone of Ser 221. It was therefore of interest to estimate the significance of Asn 155 in binding and catalysis by subtilisin. Mutating Asn 155 to Thr or His is known to have a large effect on enzyme catalytic rate⁹. We chose to simulate the Asn 155 → Ala 155 mutation for two reasons: first one could assess the function of the Asn side chain in binding and catalysis because the Ala side chain is small and has no strong interactions with its environment; and second, the simulations would be done predictively because the effect of this mutation on binding and catalysis had not been determined experimentally.

The free energy perturbation method we have used^{6,19} involves changing the molecular mechanical parameters during a molecular dynamics simulation and determining the free-energy change during this process. The basis of our study can be understood from the following scheme



This shows a simplified mechanism for enzyme action involving substrate binding E (enzyme) + S (substrate) → ES (enzyme-substrate non-covalent complex) and catalysis ES → ETS (enzyme-transition state complex). There are corresponding changes for the mutant enzyme E' and the difference in binding free energy for substrate to the two enzymes is $\Delta\Delta G_{\text{bind}} = \Delta G_{E'} - \Delta G_E$ and the difference in free energy of activation for enzyme catalysis is $\Delta\Delta G_{\text{cat}} = \Delta G_{E'}^\ddagger - \Delta G_E^\ddagger$. The processes in the scheme which involve the horizontal arrows are straightforward to measure, but difficult to study by molecular dynamics/free energy methods. As a result, the processes represented by the vertical arrows, which although not physically measurable are straightforward to simulate²⁰ are studied instead. Because free energy is a state function, the results of the computer simulations can be related to $\Delta\Delta G_{\text{cat}} = \Delta G_3 - \Delta G_2$ and $\Delta\Delta G_{\text{bind}} = \Delta G_2 - \Delta G_1$.

Thus, we must carry out three simulations; one on the enzyme itself, one on the enzyme-substrate non-covalent complex and one on a model for the transition state for acylation catalysis by the serine proteases (Fig. 1). In each case, we began with the native crystal structure of subtilisin. For ES , we dock the substrate into the active site and orient it so that its scissile bond is in position to be attacked by Ser 221, which was hydrogen-bonded to His 64. The substrate, the tripeptide Ala-Ala-Phe with neutral N and C terminals, was chosen on the basis of kinetic studies of complexes between subtilisin and a number of oligopeptide substrates¹⁴. For the enzyme in ETS , we used a structure in which His 64 was protonated and the O^- of Ser 221 formed a covalent bond with the carbonyl carbon of the substrate, which was sp^3 hybridized. We have used the charge distribution and structure of the gas phase tetrahedral complex in the hydrolysis reaction of formamide by hydroxide ion²¹ to represent the transition state complex. It must be pointed out that we have assumed that the geometrical and electronic structures of the transition state are similar to those of the gas phase tetrahedral intermediate with approximately sp^3 bonding around the scissile carbonyl carbon. We feel that these assumptions are reasonable, because of the Hammond postulate²², and also the crystallographic studies of serine proteases covalently complexed with anionic inhibitors^{17,23}. However, such types of simple model may not be appropriate for some enzymatic reactions. The molecular mechanical parameters (partial charges and so on) for this transition state model were those described by Weiner *et al.*²⁴. To change Asn 155 (system A) to Ala 155 (system B), the atoms and associated charges, van der Waals parameters, and bond length/angle/dihedral parameters for

* Present addresses: Department of Molecular Biology, Scripps Clinic, La Jolla, California 92037, USA (U.C.S.); Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138, USA (P.A.B.).

Table 1 Calculated free energy changes in the reaction illustrated in scheme (1)

Energy	ΔG_1	ΔG_2	ΔG_3	$\Delta\Delta G_{\text{bind}}$	$\Delta\Delta G_{\text{cat}}$
Electrostatic	4.59 ± 0.09	4.54 ± 0.74	7.95 ± 0.28	-0.05 ± 0.83	3.41 ± 1.02
Van der Waals	1.33 ± 0.07	1.48 ± 0.10	1.47 ± 0.21	0.15 ± 0.17	-0.01 ± 0.31
Total	5.91 ± 0.16	6.02 ± 0.64	9.42 ± 0.49	0.11 ± 0.80	3.40 ± 1.13
Experimental*				0.41	3.67

For notation, see scheme (1). Free energies are in kilocalories (1 kcal = 4.18 kJ). In all three model-built structures, the native enzyme, ES and ETS, TIP3P (ref. 34) waters were added to the system to fill the space within 18 Å of the Ser 221 O^γ using the AMBER edit program¹⁹. This was done by superimposing a cube of 5,832 water molecules (27 times the cube of 216 waters generated in Monte Carlo simulations³⁴) with its geometrical centre coinciding with the atomic coordinate of Ser 221 O^γ, and eliminating those waters which were in short contact with any atom on the protein and which were beyond 18 Å of the O^γ. The systems were energy minimized and then equilibrated for 4 ps with molecular dynamics. Then the free energy calculations were started. In all the theoretical simulations, a non-bonded cutoff of 10 Å was used. In ES and ETS, only the substrate and groups of the enzyme residues lying within 12 Å of the carbonyl oxygen of the scissile peptide were allowed to move in the molecular mechanics, molecular dynamics and free energy perturbation calculations. In the simulations on the enzyme alone (with no substrate) solvated in the active site, as earlier, only those residues within 12 Å of the Ser 221 O^γ were allowed to move. Electrostatic decoupling was employed in the free energy calculations, in that the electrostatic free energy was evaluated before the evaluation of the van der Waals energy. For the van der Waals energies, smoothing was used, to facilitate evaluation of the van der Waals free energies. In this procedure, only the van der Waals well depth changes enter explicitly into the free energy change. Comparison of two changes that involve similar changes in size of the group, this approximation seems well justified^{6,19} and the similar values for the van der Waals changes for ΔG_1 , ΔG_2 and ΔG_3 support this approximation. In the free-energy calculations, at the beginning of the simulations, each of the three structures was equilibrated at 300 K over 2,000 steps at an interval of 0.0005 ps, after which perturbations were applied with 500 steps of equilibration and 500 steps of data collection, at 0.001 ps interval. Both forward (λ from 1 to 0) and backward (λ from 0 to 1) changes in free energies were evaluated and their averages taken. In the cases, where hysteresis yielded a difference of more than 1.5 kcal mol⁻¹, the forward and backward perturbations were repeated until the above difference was less than 1.5 kcal mol⁻¹. In all the calculations, λ was varied by 0.05 at each window of equilibration and data collection. The standard deviations reported for ΔG_1 , ΔG_2 and ΔG_3 correspond to the averages of independent simulations.

* J. A. Wells and D. A. Estell (personal communication) have found that for the substrate described in ref. 5, the Ala 155 mutant has kinetic constants $K_M = 3.0 \times 10^{-4}$ M and $k_{\text{cat}} = 0.11$ s⁻¹ compared to $K_M = 1.4 \times 10^{-4}$ M and $k_{\text{cat}} = 50.0$ s⁻¹ for the native protein. This leads to the free energy changes noted here.

system A were changed to those of system B and the free energy evaluated using the following equation.

$$\Delta\Delta G = \sum_{i=1}^N \Delta\Delta G_i$$

$$= -kT \sum_{i=0}^N \ln \{ \exp[-\beta(U(\lambda_{i+1}) - U(\lambda_i))] \}_i \quad (2)$$

(For computational details see Table 1 and ref. 5.)

The results of the calculations are shown in Table 1. The calculated $\Delta\Delta G_{\text{bind}} = 0.1 \pm 0.8$ kcal mol⁻¹ and the $\Delta\Delta G_{\text{cat}} = 3.4 \pm 1.1$ kcal mol⁻¹ are qualitatively consistent with the measured quantities ($\Delta\Delta G_{\text{bind}} = 0.41$ kcal mol⁻¹ and $\Delta\Delta G_{\text{cat}} = 3.67$ kcal mol⁻¹, J. A. Wells and D. A. Estell, personal communication). This agreement shows that the critical function of the oxyanion hole in enzyme catalysis is an electrostatic effect and, furthermore, can be simulated with simple molecular mechanical models, which include only electrostatic, dispersion and exchange repulsion terms. In such a model, hydrogen bond attraction comes from electrostatic interactions.

One of the interesting conformational characteristics in the model built, and in energy-refined ES and ETS complexes involves the hydroxyl group of Thr 220. In the former, it is involved in stabilizing hydrogen bonding interactions with the

C^γ carbonyl of Asn 155. In the tetrahedral intermediate, this group swings around into the binding pocket, to form hydrogen bonding interactions with the O^{δ-} (Fig. 2). The simulations thus suggest that various double mutants of Thr 220 and Asn 155 might be of interest in studies of enzyme binding and catalysis.

Warshel and Sussman²⁵ have simulated the effect of a Gly 216, Gly 226 → Ala 216, Ala 226 mutation on $\Delta\Delta G_{\text{cat}}$ in trypsin catalysis and have found reasonable agreement with the experimental $\Delta\Delta G_{\text{cat}}$. The model used here is simpler than theirs and can be used in a straightforward way to simulate both $\Delta\Delta G_{\text{bind}}$ and $\Delta\Delta G_{\text{cat}}$. This study is the first application to both on the same system, but others²⁶⁻²⁸ have shown that the approach used here can be effective in calculating quantitatively differential free energies for binding interactions.

It is also clear from Fig. 2 that Asn 155 selectively stabilizes the transition state for catalysis because of the H-bond formed with the O^{δ-} in the transition-state model. This, of course, is not possible in the alanine mutant. Thus, it is clear that electrostatic/H-bond have important effects on enzyme catalysis, as noted previously²⁵. However, our results do not suggest that direct enzyme-transition-state interaction energies are stronger than the water-transition-state interaction energies. For example, in the solvation of ions like Na⁺, the average water-ion interaction energy is of the order of -200 kcal mol⁻¹ but the

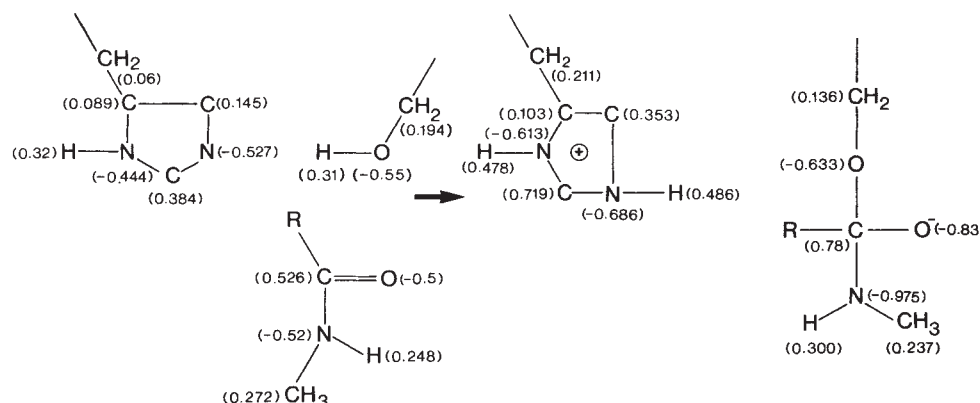


Fig. 1 Active-site charges for Michaelis and tetrahedral intermediate complexes. R, rest of the substrate. The remaining charges not shown in the figure are given in ref. 25. This simulation used a united atom model for hydrogens bonded to carbon.

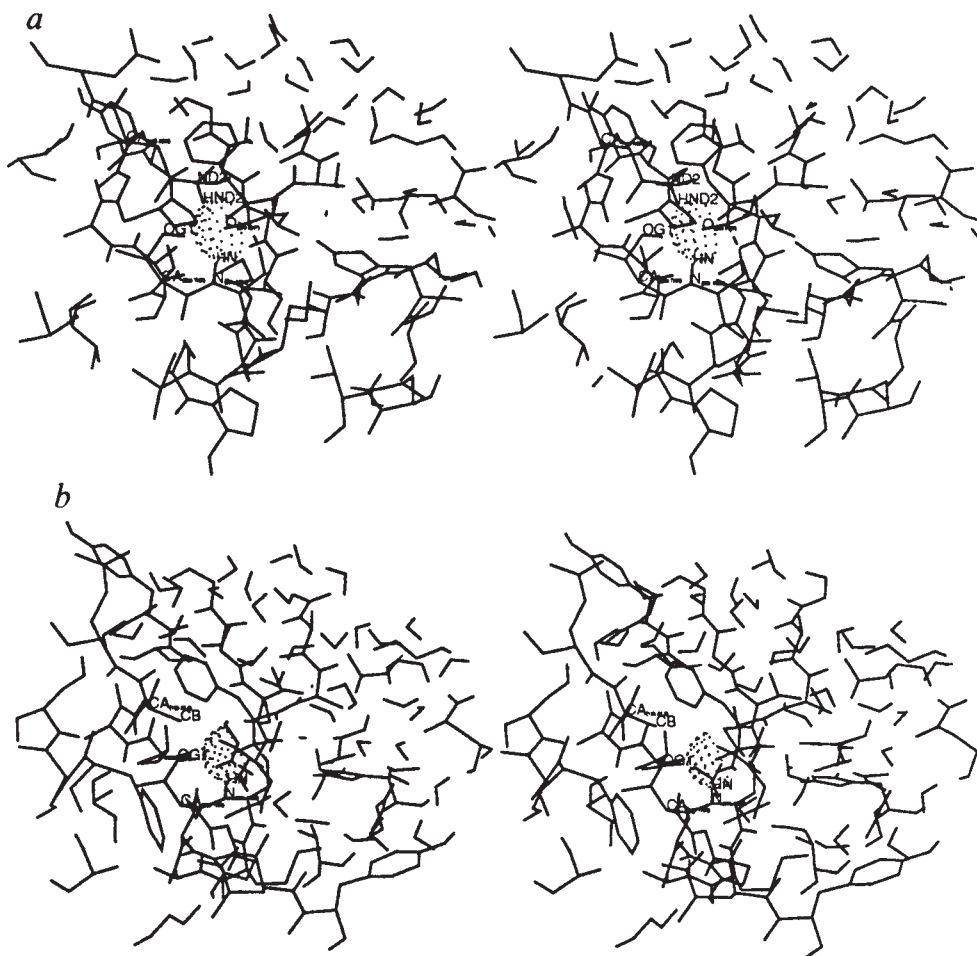


Fig. 2 Stereo view of the active-site region of the model for transition state/tetrahedral intermediate of *a*, native enzyme (obtained after energy minimization of the model-built complex and subsequent equilibration using molecular dynamics); and *b*, Ala-155 mutant (obtained after the perturbation of Asn 155 to alanine as described in the text). Residues lying within 10 Å of the scissile peptide carbonyl carbon are shown. The oxygen of the carbonyl in the scissile peptide bond is shown with its van der Waals surface. Atoms which are involved in hydrogen bonding interactions with this oxygen are labelled. They are N2 and HND2 in Asn 155, atoms in the hydroxyl group of Thr 220 and atoms in the backbone N—H of Ser 221. In *b*, residue 155 is an alanine and hence only the CB atom is labelled. No water molecule moved into the hole created as a result of the Asn to Ala mutation, even though one of the water molecules surrounding the active site could have done so. Either kinetic effects (the simulations only lasted tens of picoseconds) or the intrinsic instability of a water molecule in this pocket may have prevented such a water movement. Bash³⁵ did find that a water molecule moved into the hole created by an Asn to Ser mutation in dihydrofolate reductase during his simulations.

unfavourable water–water interaction energy is +100 kcal mol⁻¹, thus making the net energy of interaction -100 kcal mol⁻¹ (ref. 29). A macromolecule or crown ether³⁰ with its pre-aligned dipoles can have a macromolecule–Na⁺ interaction energy of only -105 kcal mol⁻¹ and still have a more favourable net interaction energy than the net water–ion interaction energy. The same argument applies to free energies and explains how the macromolecular oxyanion hole, with two reasonable N—H...O hydrogen bonding interactions, can stabilize the transition state for peptide hydrolysis better than can water, even with its three very favourable O—H...O hydrogen bonding interactions²¹. In the enzyme catalysed reaction, the enzyme has paid a free-energy price during synthesis to line up its oxyanion hole N—H bonds in the same direction and thus in an electrostatically unfavourable orientation, which, however, allows these N—H bonds to interact well with the substrate oxyanion. It has paid that free energy price because other favourable free-energy terms such as burial of hydrophobic groups, compensate for such unfavourable N—H alignments. Support for this comes from the calculated ΔG_1 of 5.9 kcal mol⁻¹, which is significantly smaller than the value found⁵ for the difference in solvation free energy of acetamide and methane. Thus, in our opinion, the most important feature that differentiates an enzyme catalysed reaction from the corresponding solution reaction is not electrostatics in themselves, but the free energy price paid in synthesis to align^{31,32} and desolvate³³ the catalytic functional groups.

We emphasize that the method as implemented here will only be successful if the structural perturbation due to the mutation is rather localized and, thus, can be simulated in the 10–100-ps time range. The ability to simulate $\Delta\Delta G_{\text{cat}}$ depends on a reasonable structural model for the transition state for catalysis, and the agreement with experiment supports the notion that our model is reasonable, although this cannot be proved. We note

that the serine proteases are also a good choice for study because the evidence for the tetrahedral nature of the transition state for acylation is strong. But the simulation method described here should be useful when combined with site-specific mutagenesis effects for determining the transition state for catalysis when experimental studies are more equivocal than in the serine proteases.

We thank Drs R. Bott, A. Kossiakoff and B. Katz of Genetech for providing the coordinates for native subtilisin used here and to Drs J. Wells and D. A. Estell for unpublished data on the mutant provided after our simulations were completed. NIH support to P.A.K. and to the UCSF Computer Graphics Laboratory is gratefully acknowledged, as is Cray X-MP computer time provided at the San Diego Supercomputer Center by an NSF grant. The FPS-264 used for some of the calculations was purchased through grants from the NSF and NIH. We thank Merck, Sharp and Dohme for support.

Note added in proof: Hwang and Warshel³⁶ have independently calculated the catalytic free energies and binding free energies of native and mutant subtilisins in which Asn 155 was replaced by Ala, Leu and Thr, and found results consistent with ours.

Received 11 February; accepted 29 May 1987.

1. Wells, T. N. C. & Fersht, A. R. *Nature* **316**, 656–657 (1985).
2. Strauss, D., Raines, R., Kawashima, E., Knowles, J. R. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **82**, 2272–2276 (1985).
3. Craik, C. S. *et al. Science* **228**, 291–297 (1985).
4. Kollman, P. A. *Acc. Chem. Res.* **18**, 105–111 (1985).
5. Bash, P. A., Singh, U. C., Langridge, R. & Kollman, P. A. *Science* **236**, 564–568 (1987).
6. Singh, U. C., Brown, F. K., Bash, P. A. & Kollman, P. A. *J. Am. Chem. Soc.* **109**, 1607–1614 (1987).
7. Robertus, J. D., Kraut, J., Alden, R. A. & Birktoft, J. J. *Biochemistry* **11**, 4293–4303 (1972).
8. Poulos, T. L., Alden, R. A., Freer, S. T., Birktoft, J. J. & Kraut, J. *J. biol. Chem.* **250**, 1097–1103 (1976).
9. Wells, J. A., Cunningham, B. C., Graycar, T. P. & Estell, D. A. *Phil. Trans. R. Soc. A317*, 415–423 (1986).
10. Estell, D. A. *et al. Science* **233**, 659–663 (1986).

11. Estell, D. A., Graycar, T. P. & Wells, J. A. *J. biol. Chem.* **260**, 6518-6521 (1985).
12. Wells, J. A. & Powers, D. B. *J. biol. Chem.* **261**, 6564-6570 (1986).
13. Power, S. D., Adams, R. M. & Wells, J. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3096-4000 (1986).
14. Robertus, J. D. *et al. Biochemistry* **11**, 2439-2449 (1972).
15. Katz, B. A. & Kossiakoff, A. *J. biol. Chem.* **261**, 15480-15485 (1986).
16. Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. *Nature* **214**, 652-656 (1967).
17. Stroud, R. M., Kay, L. M. & Dickerson, R. E. *J. molec. Biol.* **83**, 185-208 (1974).
18. Shotton, D. M. & Watson, H. C. *Nature* **225**, 811-816 (1970).
19. Singh, U. C., Weiner, P. K., Caldwell, J. W. & Kollman, P. A. *AMBER(UCSF) Version 3.0* (Dep. Pharmac. Chem., Univ. California, San Francisco, 1986).
20. Tembe, B. L. & McCammon, J. A. *J. comput. Chem.* **8**, 281-284 (1984).
21. Weiner, S. J., Singh, U. C. & Kollman, P. A. *J. Am. chem. Soc.* **107**, 2219-2229 (1985).
22. Hammond, G. S. *J. Am. chem. Soc.* **77**, 334-338 (1955).
23. Kossiakoff, A. A. & Spencer, S. A. *Nature* **288**, 414-416 (1980).
24. Weiner, S. J. *et al. J. Am. chem. Soc.* **106**, 765-784 (1983).
25. Warshel, A. & Sussman, F. *Proc. natn. Acad. Sci. U.S.A.* **83**, 3806-3810 (1986).
26. Lybrand, T., McCammon, J. A. & Wipf, G. W. *Proc. natn. Acad. Sci. U.S.A.* **83**, 833-836 (1986).
27. Wong, C. F. & McCammon, J. A. *J. Am. chem. Soc.* **108**, 3830-3832 (1986).
28. Bash, P., Singh, U. C., Brown, F., Langridge, R. & Kollman, P. *Science* **235**, 574-576 (1987).
29. Chandrasekhar, J., Spellmeyer, D. & Jorgensen, W. L. *J. Am. chem. Soc.* **106**, 903-910 (1984).
30. Wipf, G. W., Weiner, P. K. & Kollman, P. A. *J. Am. chem. Soc.* **104**, 3249-3258 (1982).
31. Warshel, A. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5250-5254 (1978).
32. Warshel, A. *Biochemistry* **20**, 3167-3177 (1981).
33. Weiner, S. J., Seibel, G. L. & Kollman, P. A. *Proc. natn. Acad. Sci. U.S.A.* **83**, 649-653 (1986).
34. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein M. L. *J. chem. Phys.* **79**, 926-935 (1983).
35. Bash, P. A. thesis, Univ. California, Berkeley (1986).
36. Hwang, J.-K. & Warshel, A. *Biochemistry* **26**, 2669-2673 (1987).

Evidence for kinks in DNA folding in the nucleosome

Michael E. Hogan*, Thomas F. Rooney* & Robert H. Austin†

* Department of Molecular Biology and † Department of Physics, Princeton University, Princeton, New Jersey 08544, USA

The nucleosome subunit of chromatin consists of DNA folded around a histone core as a 1.8-turn left-handed solenoid. The crystal structure¹ of the nucleosome core particle revealed that it has a dyad symmetry axis and that the minor helix groove faces outwards from the protein core. Richmond *et al.*¹ noticed that the path traversed by the helix has severe bends at sites approximately one and four helix turns from the dyad axis. We have developed two photochemical methods to study the structure of DNA, and in particular that wrapped around the nucleosome core. One method depends on the sensitization of singlet oxygen production by an eosin analogue. We have monitored the rate at which excited state oxygen diffuses into contact with DNA base planes, and find that it attacks the nucleosome with high specificity. We have also mapped the DNA binding of the intercalating dye methylene blue, and conclude that it binds to the same sites accessible to oxygen by diffusion. On the basis of these results we suggest that the DNA in the nucleosome is bent or kinked at two sites, 1.5 helix turns from the dyad axis.

To extend our understanding of variation in DNA structure in the nucleosome and other large complexes, we have exploited a photochemical method based on the use of eosin in covalent complex with tris hydroxyamino methane (ET). The eosin ring system is negatively charged and for that reason does not bind to DNA. When complexed to form the zwitterion, the dye is not expected to bind protein strongly either. Therefore as expected, we have been unable to detect spectroscopically ET binding to nucleosomes or to purified 146-base pair (bp) DNA at concentrations as high as 10^{-3} M base pairs (not shown).

When irradiated in the presence of dissolved molecular oxygen, ET sensitizes the production of singlet oxygen (O_2^1) which will oxidize purine upon diffusing into collisional contact with the helix. The altered purine is a stable adduct. However, when heated, the modified helix is broken at oxidized purine sites, by means of an elimination reaction similar to that initiated by alkylation². This dye-mediated photochemistry is not sensitive to DNA sequence variation in itself (the rate of strand breakage

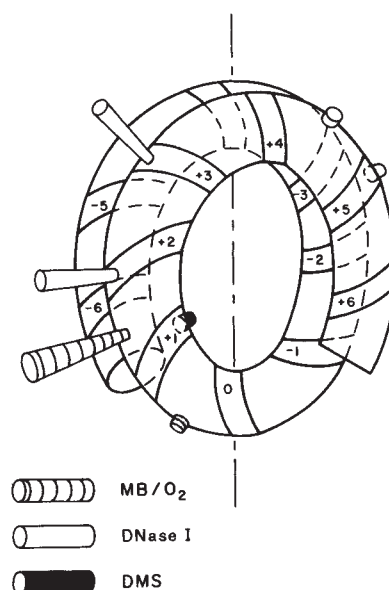


Fig. 1 Map of DNA cleavage sites on the nucleosome. The 146-bp DNA complement of a nucleosome is presented as 1.8 turns of left-hand solenoid. The overall dimensions of the solenoid and the disposition of helix grooves are as specified by the 7-Å crystal structure (grooves are represented by transverse banding). No attempt has been made to portray additional fine structure. Methylene blue and O_2^1 attack sites, and DNase I (additional data from ref. 10) and DMS¹¹) cleavage sites have been marked by cylinders of length proportional to the cleavage rate at each site (unmarked positions correspond to relative rates <5% of the maximum). A diad axis has been superimposed on this low resolution representation of the nucleosome, passing outward through the minor helix groove at base pair 73, as specified by crystallography¹. Numbering of groove positions (0, 1, 2, and so on) corresponds to the nomenclature of ref. 1).

at A or G is independent of sequence). Therefore, this photochemistry has been shown to be a useful method for DNA sequence analysis². By employing an intense argon ion laser, we have modified that sequencing procedure for mapping O_2^1 attack sites on a DNA, at one base resolution.

In an earlier study, we showed that if contiguous base pairs in a helix are modelled as forming a wedge-shaped cavity, then the rate of oxygen diffusion into such a site will be proportional to the opening angle of the wedge³. If O_2^1 -mediated strand break chemistry is limited by such diffusional barriers, then the distribution of cleavage rates in a DNA region will be proportional to local variation of base stacking (see Fig. 2 legend for details).

To test that proposal, we have compared ET-mediated cleavage rates for native and heat denatured DNA (Fig. 2b and c). When duplex DNA is boiled and then quenched on ice before ET treatment, the apparent first-order O_2^1 attack rate of denatured DNA increases by a factor of 21 ± 3 .

Heat denatured DNA in these measurements is 78% single stranded (see Fig. 2 legend). Assuming an opening angle equal to 180° for an unstacked base plane, and a root mean square (r.m.s.) bending angle between base pairs in native DNA equal to 5° (ref. 4) these rate arguments predict that heat denatured DNA should be cleaved $[(0.78) \times (180/5) + (0.22) \times (5/5)] = 24$ times faster than an intact duplex. Considering the geometric simplifications inherent in this rate analysis, we feel that the measured 21-fold rate difference is in good agreement with the model. From that agreement, we conclude that O_2^1 attack rates are diffusion limited in DNA, and as such can be used to measure base stacking variation within a helix.

To map the position of O_2^1 attack sites, nucleosomes were treated with ET, irradiated at 514 nm, deproteinized, 5' end-labelled with ^{32}P , treated with piperidine to induce hydrolysis

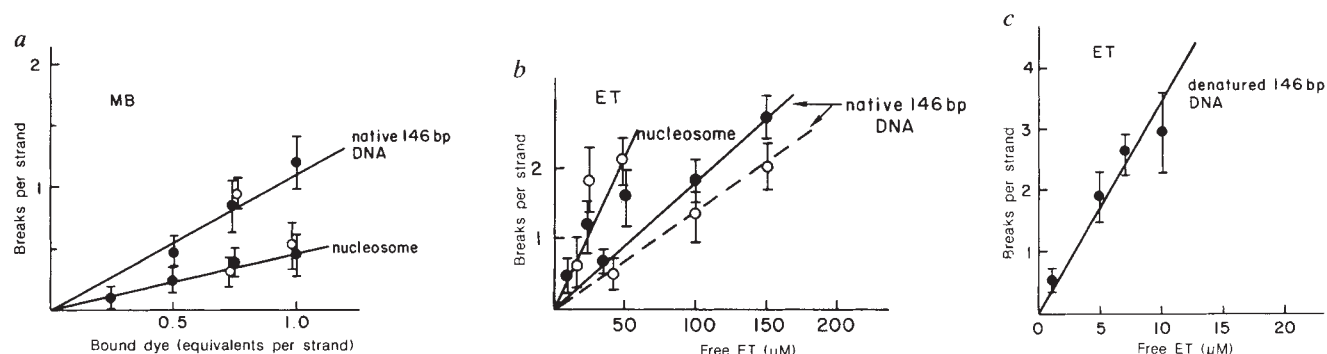


Fig. 2 Photochemical cleavage of DNA, using methylene blue (a), or eosin-Tris complex (b, c). Error bars, one standard deviation about the mean of five measurements. a, b, Cleavage of native 146-bp DNA and nucleosomes. ●, Results with Tris buffer alone; ○, results with Tris buffer plus 650 μM CaCl₂. c, Cleavage of denatured 146-bp DNA.

Methods. Nucleosome core particles were prepared by limited micrococcal nuclease digestion of chicken erythrocyte nuclei, and 146-bp DNA fragments were prepared from nucleosomes by proteinase K treatment⁶. In all instances, sample purity was confirmed by protein and DNA electrophoresis. Nucleosomes and 146-bp DNA fragments were 5' endlabelled with T4 polynucleotide kinase (Boehringer Mannheim) and ATP (ref. 12), then repurified by chromatography over Sephadex G50. Samples were concentrated by pressure filtration on a Centricon 30 system (Millipore). Denatured DNA samples were prepared by mixing 5' labelled 146-bp DNA (1×10^{-8} M bp (1×10^{-8} M in base pairs)) with unlabelled sheared calf thymus DNA (5×10^{-4} M bp (5×10^{-4} M in base pairs)) both in 1 mM Tris-HCl, 10 μM EDTA. The mixture was boiled for 3 min, then quenched on ice. We have determined that after this treatment, DNA absorbance at 260 nm increases by $31 \pm 3\%$. Assuming that complete denaturation is accompanied by a 40% absorbance increase, we calculate that $(1 - 31/40) \times 100 = 22\%$ of base pairs remain in a duplex state after the heat-quench treatment, the extent of renaturation expected from other studies. Native 146-bp fragment samples were obtained identically except that the boiling step was omitted. Methylene blue (MB) was purchased from Eastman Kodak and repurified by LH20 chromatography⁶. Concentration was determined setting $\epsilon_{665} = 8 \times 10^4 \text{ M}^{-1} \text{ cm}^{-1}$. The eosin-Tris complex (ET) was prepared by reacting eosin 5-isothiocyanate (Molecular Probes) with 10 mM Tris-HCl, pH 7.8 at 37 °C for 2 h. The reaction progressed to completion, as assessed by TLC. For MB reactions, MB was added to DNA or nucleosome samples to the desired bound density. We have determined that MB binding to DNA and nucleosomes is stoichiometric in this buffer at 5×10^{-4} M bp (ref. 7). Irradiation was performed with a HeNe laser operating at 250 mW cm^{-2} at 632.8 nm. Irradiation time was constant at 15 min. Nucleosome samples were then deproteinized, ethanol precipitated, resuspended in 10% piperidine, and heated for 30 min at 90 °C, to induce DNA hydrolysis at $\text{O}_2^{\bullet}$ oxidation sites, as described by Friedman and Brown². The 146-bp DNA samples were treated with piperidine directly. After hydrolysis, DNA was vacuum dried, resuspended in deionized formamide, boiled for 3 min, then electrophoresed over a standard 8% sequencing gel. The fraction (*f*) of DNA strands which remained uncleaved during a photochemical treatment was quantified from autoradiograms, using a BioRad 620 diode array densitometer. The average number of photochemically-induced breaks per strand (*n*) was calculated assuming Poisson statistics from $n = -\log_e(f)$. In all instances, the cleavage data were adequately fitted as first order with respect to dye concentration. The corresponding linear fits to the data have been generated by a least-squares analysis. For the technique using the eosin-Tris complex, ET was added to the desired concentration, then irradiated with an argon ion laser at 2.5 W cm^{-2} at 514 nm. Irradiation time was constant at 3 min. Sample treatment after irradiation was as for MB. As discussed in the text, if $\text{O}_2^{\bullet}$ attack of DNA is diffusion limited, then local variation of purine oxidation will be proportional to base attacking variation in a helix region. If R_i/R_j is the relative attack rate at two sites, then $R_i/R_j = [(\theta_i^2/\theta_j^2)]^{1/2}$ where $(\theta_i^2)^{1/2}$ is the r.m.s. angle formed by the stacking of adjacent base planes. (See ref. 3 for details). With denatured DNA, ET was added directly to heat denatured DNA, then irradiated as above.

at oxidized purine, then displayed on a standard sequencing gel.

We find that a single site is attacked very rapidly by $\text{O}_2^{\bullet}$ to produce a DNA fragment of 88 ± 3 bases (Fig. 3a). From densitometry of the data, we conclude that the $\text{O}_2^{\bullet}$ attack rate there is 20 times that of other sites in the nucleosome. As expected from the randomness of DNA sequence in these nucleosomes, this pronounced site specificity is not detected with purified 146-bp DNA fragments (data not shown).

It is useful to compare $\text{O}_2^{\bullet}$ attack to the cleavage produced by DNaseI. As evident from Fig. 3b, c, ET-mediated cleavage is displaced from DNaseI cleavage sites by an average of four bases. Because DNaseI cleaves the nucleosome where the minor helix groove faces outward¹ our data show that $\text{O}_2^{\bullet}$ attack occurs where the major helix groove is outward facing. In the context of the low-resolution crystallography, we conclude that the principle site of $\text{O}_2^{\bullet}$ attack occurs at a pair of symmetry related sites, 1.5 helical turns from the particle dyad. The position of that site is shown in Fig. 1, superimposed on a cartoon of nucleosome structure.

Figure 2b describes an important control experiment which shows that the overall rate of $\text{O}_2^{\bullet}$ attack on the nucleosome is 2.2 times that on purified DNA. That enhancement suggests that $\text{O}_2^{\bullet}$ cleavage specificity at 1.5 results (following the notation of ref. 1) from a base stacking change there, rather than to a general reduction of cleavage at other sites in the particle. For that reason, it is appropriate to apply the geometric model discussed above and in ref. 3, from which we calculated that, if modelled

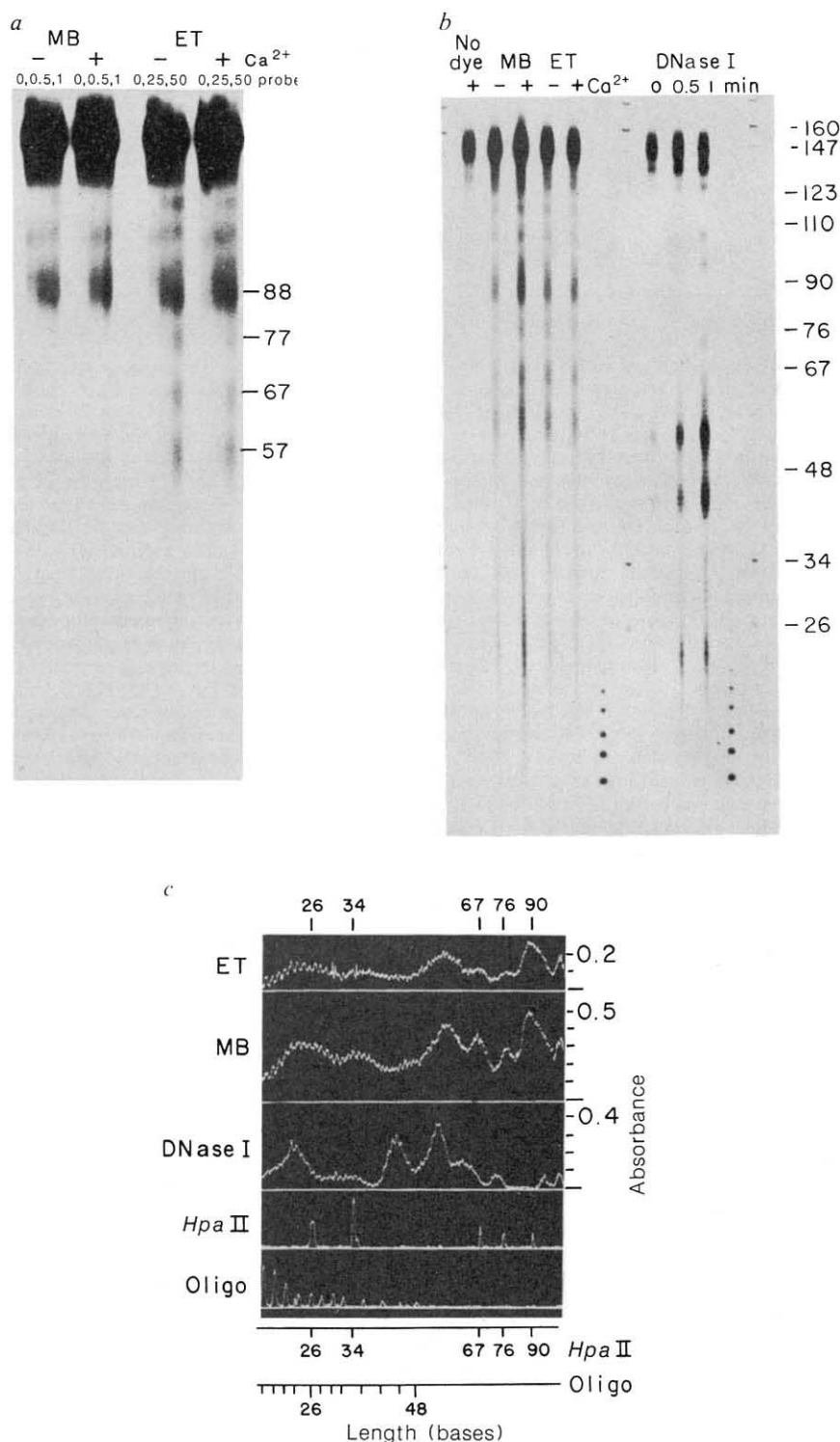
as a wedge, the average base opening angle at site 1.5 may be 20 times greater than that of a linear helix segment.

That kind of base stacking decrease could result from a sharp bend or kink in the helix fold. Richmond and colleagues have identified such an elbow in their 7 Å refinement of nucleosomes structure, near site 1 (ref. 1). It is likely that the elbow in the crystal structure is related to the site which we have mapped as an $\text{O}_2^{\bullet}$ attack site. However, our data require that the bend occurs 0.5 helix turns away from the site deduced from the low-resolution crystal structure. Additional work is required to reconcile that apparent difference.

Richmond *et al.*¹ concluded that at a low resolution, the DNA bend near 1 does not express similarity to any kinked helix structure which has been identified. On the other hand, we have found it useful to consider our $\text{O}_2^{\bullet}$ mapping data in relation to a model for DNA bending first proposed by Crick and Klug⁵, which involves opening of base planes toward the major helix groove, and as such makes two predictions: First, if a helix in the nucleosome is to remain folded, such a kink must be positioned where the major helix groove faces outward from the particle. Second, adjacent bases at such a kink form a wedge-shaped cleft, where the average base stacking angle has increased to 90°, from a time average value of 5° in a linear helix (approximately an 18-fold increase). We feel that our $\text{O}_2^{\bullet}$ mapping data are in good agreement with both aspects of that model.

In order to explore that DNA structure variation in more detail, we have exploited a second photochemical method, based

Fig. 3 High-resolution mapping of MB and ET photochemistry. DNA and nucleosome samples were prepared then irradiated as in Fig. 2. Photochemical cleavage sites have been measured by high-resolution electrophoresis on a 7 M urea, 8% acrylamide, 0.2% bisacrylamide gel. Two classes of size standard were employed: a synthetic oligonucleotide marker set, labelled with ATP and polynucleotide kinase (representing 48, 44, 40, 36, 32, 30, 28, 26, 24, 22, 20, 18, 16, 14, 12 bases) and the Klenow-labelled *Hpa*II digest of pBR322, which produces, among others, DNA fragments of 123, 110, 90, 76, 67, 34, 26 bases. In the region of overlap, the two sets of standards were found to agree to within one base accuracy. *a*, Mapping on nucleosome. Nucleosomes were treated with 0, 0.5 or 1 bound MB equivalent per strand, then irradiated as in Fig. 2. ET-mediated (O_2^*) cleavage was performed by adding ET to 0, 25 or 50 μ M, then irradiation as in Fig. 2. Irradiations were also performed on samples in the standard Tris buffer plus 650 μ M total $CaCl_2$, corresponding to 400 μ M free Ca^{2+} , as determined from the Ca^{2+} binding isotherm in ref. 7. Size markers are in bases. *b*, Comparison of photochemical and DNaseI cleavage. The ^{32}P -labelled nucleosomes were digested with DNaseI in 10 mM Tris-HCl, 650 μ M $CaCl_2$, 20 $^\circ$ C, at 20 units ml^{-1} . Nucleosome concentration was 5×10^{-4} M bp. The data correspond to 0, 0.5 or 1 min digestion, quenched by EDTA and phenol extraction. For comparison, MB-nucleosome complexes were irradiated at 2 dye equivalents per strand, to emphasize primary and secondary binding sites. ET was administered at 200 μ M; again, to measure O_2^* cleavage at primary and secondary attack sites. *c*, Densitometry of MB, ET and DNaseI analysis. Autoradiograms as in *b* were analysed by microdensitometry. Data are presented as optical density plotted against position. Top, ET-treated nucleosomes, 200 μ M ET, no Ca^{2+} . Second row, MB-treated nucleosomes, 2 MB equivalents per strand, no Ca^{2+} . Third row, DNaseI-treated nucleosomes, 0.5 min incubation. Bottom two rows: *Hpa*II-digest and oligonucleotide size markers; the horizontal axes have been marked with the position of appropriate length standards.



upon the use of methylene blue (MB) as a structure probe. MB is a well characterized DNA intercalator which we have used as a spectroscopic probe of DNA structure and dynamics^{6,7}. As for ET, MB can sensitize the production of O_2^* (ref. 2). Because the local concentration of an unstable molecule such as O_2^* drops rapidly with distance from a bound point source (with a dependence at least as steep as r^{-3}) base oxidation should be localized to the MB binding site. Consequently, the distribution of strand breaks which results from this method should be representative of the distribution of bound dye molecules.

Sobell⁸ and others⁹ have proposed that transient formation of a kinked or unstacked secondary structure may limit the kinetics of intercalator binding. A corollary of that proposal is that a preformed bend or kink might attract the binding of MB.

If such kinks form in the nucleosome, MB binding could be targeted to those regions.

Under conditions where >95% of the dye is bound by intercalation⁷ we find that MB-mediated strand breaks are produced in the nucleosome at a rate which is 0.6 times that measured for uncomplexed DNA (Fig. 2a). To map the position of those

cleavage sites, nucleosome particles have been modified with MB, irradiated at 632 nm, then analysed as described for ET.

At the lowest bound density, MB cleaves the nucleosome specifically, at a site 88 ± 3 bases from the 5' DNA terminus (Fig. 3a, leftmost). When the bound dye is increased to two molecules per strand, cleavage also becomes measurable at other sites (67 and 77 bases, Fig. 3b) which may be indicative of binding saturation at the principle cleavage site. Again, site specific cleavage is not detected when these experiments are repeated on purified 146-bp DNA fragments (not shown). The simplest model which is consistent with these data is that MB binds specifically, then cleaves the open helix structure which we have detected at site 1.5. Work is in progress to confirm this interpretation of the MB mapping data, using photochemical dye probes which cleave DNA stoichiometrically.

We have shown that DNA structure in the nucleosome is altered by multivalent ion binding. Most notably, base stacking in the particle may be altered (as monitored by circular dichroism) and the properties of the MB binding site change⁷.

To explore photochemically how ion binding alters DNA structure, we undertook MB- and ET-mediated O_2^* mapping experiments in the presence of 400 μ M free Ca^{2+} . It is obvious from these data (Fig. 3) that MB- or ET-mediated attack does not change significantly upon binding Ca^{2+} . Yet, biophysical analysis has shown that the local structure of the MB binding site is altered by ion binding⁷. It remains to be determined how those spectroscopic changes are related to DNA structure in the bent region at 1.5.

The work described here represents the first application of a useful set of photochemical tools for DNA structure analysis. Using MB as a probe, we have shown that it is possible to monitor the position of intercalator binding sites on a protein-DNA complex at one to two base resolution. We believe that data of that kind can be generally useful for the following reasons. First, the mechanism of intercalator binding is well known and interpretable in molecular detail. Second, the photochemistry sensitized by MB is efficient enough for one-base-resolution mapping to be performed with one to two added equivalents per helix at relatively low light levels. Third, because MB is a vital stain, and because MB photochemistry is efficient and easy to control, it is likely that the methodology can be extended for the purpose of DNA structure analysis in whole cells. Finally, MB is a powerful spectroscopic probe. As a result, if DNA or a protein-DNA complex display structure variation which attracts the dye, it is possible to investigate those high affinity sites spectroscopically.

In a second class of experiment, we have determined the position and relative rate of ET mediated O_2^* attack on DNA. In addition to being efficient, easily controlled, and potentially applicable to analysis *in vivo*, ET-mediated oxidation seems to be diffusion limited. As a result, the distribution of DNA strand breaks may be interpreted quantitatively, in terms of local variation of base stacking in a helix. Work is in progress to evaluate the full potential of this family of photochemical methods.

We acknowledge the support of the Office of Naval Research (R.H.A.), and of the National Cancer Institute and the National Science Foundation (M.E.H.). Some of this work was completed after Thomas Rooney's tragic death.

A plant processed pseudogene

Guy Drouin & Gabriel A. Dover

Department of Genetics, University of Cambridge, Cambridge CB2 3EH, UK

Processed pseudogenes are a common feature of mammalian multi-gene families and 'single-copy' genes¹⁻³. They are considered to result from integration into the genome of reverse transcripts of cytoplasmic poly(A)⁺ messenger RNA of genes that are expressed in the germ line. The paucity or absence of processed pseudogenes in animal groups other than mammals, despite the presence of endogenous viral retrotransposons encoding reverse transcriptase⁴, has led to the suggestion that this disparity might be the result of differences in the duration of gametogenesis³. In contrast to animals, plants lack a germ line: plant gametes differentiate from cells that have been actively engaged in organizing the somatic tissue of the plant, in the apical meristem. Pseudogenes of somatically expressed genes could therefore appear eventually in the gametes. Although reverse transcription is known to be involved in the replication of the cauliflower mosaic virus⁵⁻⁷, and retrotransposons have been described in wheat and maize⁸⁻¹⁰, surprisingly no plant processed pseudogene has been reported. We report here the nucleotide sequence of an actin gene from potato (*Solanum tuberosum*) which has all the characteristic features of a typical mammalian processed pseudogene. This potato processed pseudogene might be the first of many that are yet to be discovered. Alternatively, it might be a rare event because some, as yet, unknown factors affect the different rates of pseudogene production in animals and plants.

In characterizing the multiple actin genes present in the potato genome we encountered a particular actin gene which lacks the three intervening sequences characteristic of all plant actin genes so far sequenced, including five other non-allelic potato actin genes that we are currently studying (Fig. 1). The complete sequence of this gene is shown in Fig. 2. Note that the introns of this gene (their actual location in other plant actin genes is shown by triangles) have been precisely removed, suggesting that a processed RNA intermediate was probably involved in the formation of this pseudogene. The presence of an 8-basepair (bp) oligo(A) tract preceding the downstream copy of an 18-bp direct repeat is also indicative of formation through an RNA intermediate. The direct repeats indicate that there has been an integration of cDNA into a previously unoccupied position in the nuclear genome. The pseudogene is characterized by the

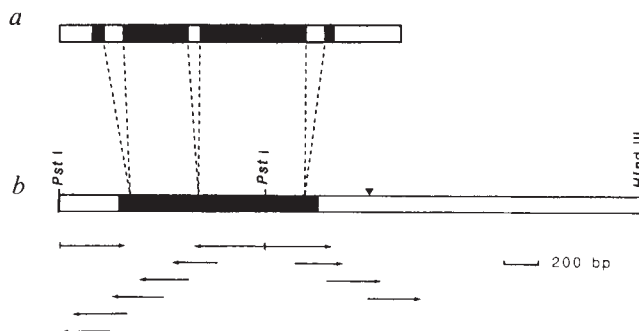


Fig. 1 a, Typical structure of a plant actin gene with four exons. b, Restriction map of the sequenced 2,082-bp region which was subcloned from an *EcoRI*-cut EMBL4 phage (M. Thangavelu *et al.*, in preparation). Arrows underneath the map, the pEMBL18 and 19 inserts¹⁴, and their *Bal31* deletion products¹⁵, indicate regions which were sequenced by the dideoxy procedure¹⁶. The coding region is filled in. Triangle, poly(A) tract.

Received 10 March; accepted 29 May 1987.

- Richmond, T. J., Finch, J. T., Rushton, B., Rhodes, D. & Klug, A. *Nature* **311**, 532-537 (1984).
- Freidman, T. & Brown, D. M. *Nucleic Acids Res.* **5**, 615-622 (1978).
- Berkoff, B., Hogan, M. E., LeGrange, J. & Austin, R. *Biopolymers* **25**, 307-316 (1986).
- Barkeley, M. D. & Zimm, B. H. *J. chem. Phys.* **70**, 2991-3007 (1979).
- Crick, F. H. C. & Klug, A. *Nature* **255**, 530-533 (1975).
- Wang, N. C., Hogan, M. E. & Austin, R. H. *Proc. natn. Acad. Sci. U.S.A.* **79**, 5896-5900 (1982).
- Hogan, M. E., Hayes, B., Wang, N. C. & Austin, R. H. *Biochemistry* **25**, 5070-5082 (1986).
- Sobell, H. M., Tsai, C. C., Gilbert, S. G., Jain, S. C. & Sakore, T. D. *Proc. natn. Acad. Sci. U.S.A.* **73**, 3068-3072 (1976).
- Li, H. J. & Crothers, D. M. *J. molec. Biol.* **39**, 461-477 (1969).
- Lutter, L. C. *Nucleic Acids Res.* **6**, 41-57 (1979).
- McGhee, J. & Felsenfeld, G. *Proc. natn. Acad. Sci. U.S.A.* **76**, 2133-2137 (1979).
- Maniatis, T., Fritsch, E. F. & Sambrook, J. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Laboratory, New York, 1982).

CTGCAGTGATTCCTATAAATCTGCTACCTAAATATAGCTATGTATATTACTAGATTTAACTCTATATACACTCTATAACATGGAGGAACCATCTACTTACAGATTGACTTCAAAT 116
 CTTTGTGATGAAATAAGACTAATTTGGAACTTCTGGAGTGGTACGAATGGACGGAAGAGGAAGTGCATCATCAGAACTTCTGTGACAGAAATCAATGAGTGGATTGTTTGA 235
 CATTGCGATAGGTACTCATTTTGTATTCTCCATTTCCCAAACTGAGTCTTCCGAGCCATGGCCATAGTCTCCGGTTGCTTTGCTAAATTTCCGGATACCAATTTCCCGATCTCCTCTC 354

+1 Met Ala Asp Gly Glu Asp Ile Gln Pro Leu Val Cys Asp Asn Gly Thr Gly Met Val Lys Ala Gly Phe Ala Gly 448
 ACTTCTGTGTAATAGAAA ATG GCT GAT GGT GAG GAC ATT CAG CCC CTT GTT TGT GAC AAT GGA ACT GGA ATG GTG AAG GCT GGA TTT GCT GGT

+30 Asp Asp Ala Pro Arg Val Val Phe Thr Ser Ile Val Gly Arg Pro Arg His Thr Gly Val Met Val Gly Met Arg Gln Lys Asp Ala Tyr 538
 GAT GAT GCT CCT AGA GTT GTG TTC ACT AGC ATT GTT GGT CGT CCT CGC CAC ACT GGT GTT ATG GTT GGA ATG AGA CAG AAG GAT GCA TAT

+60 Val Gly Asp Glu Ala Gln Ser Lys Arg Gly Ile Leu Thr Leu Leu Ile Pro Ser Glu His Gly Ile Val Ser Asn Trp Asp Asp Met Glu 629
 GTT GGT GAT GAA GCT CAA TCC AAG AGG GGT ATC TTG ACC TTG AAA ATAGCCA AGT GAG CAT GGT ATA GTC AGC AAC TGG GAT GAT ATG GAG

+90 Lys Ile ... His His Thr Phe Tyr Asn Glu Leu His Val Ala Pro Glu Glu His Pro Val Val Leu Thr Glu Ala Pro Leu Asn Pro Lys 719
 AAG ATA TGA CAT CAT TTT TAC AAT GAG CTT CAT GTT GCC CCC GAG GAG CAC CCT GTT GTG CTC ACT GAA GCA CCT CTC AAC CCT AAG

+120 Pro Asn Arg Glu Lys Met Thr Gln Ile Ile Phe Xxx Thr Phe Asn Val Pro Ala Met Thr Val Val Ala Ile Gln Ala Val Leu Ser Leu Tyr 809
 CC- AAC AGA GAG AAA ATG ACC CAG ATT ATA TTT A-G ACC TTC AAT GTT CCG GCT ATT GAT TAT GTT GCT ATT CAG GCT GTG CTT TCC TTG TAT

+150 Ala Ser Gly Arg Thr Thr Gly Ile Val Leu Asp Cys Gly Asp Gly Val Ser His Thr Val Pro Met Tyr Glu Gly Tyr Leu Xxx Pro His 899
 GCT AGT GGT CGT ACA GGT ATT GTG TTG GAG TGT GGT Cys Gly Asp Gly Val Ser His Thr Val Pro Met Tyr Glu Gly Tyr Leu Xxx Pro His

+180 Ala Ile Leu Arg Leu Asp Ser Gly Arg Asp Pro Thr Asp Asn Leu Met Lys Ile Leu Thr Glu Arg Gly Tyr Met Phe Thr Thr Thr 989
 GCC ATT CTT CGT TTG GAT CTT TCT GGC CGT GAT CCA ACT GAT AAC CTG ATG AAG ATT CTC ACT GAG AGA GGT TAC ATG TTC ACC ACC ACT

+210 Ala Glu Arg ... Ile Val Cys Asp Met Lys Val Lys Leu Ala Tyr Val Ala Leu Asp Tyr Glu Gln Glu Met Xxx Thr Ala Arg Ser Ser 1079
 GCT GAG CGG TAA ATT GTC TGT GAC ATG AAG GTA AAA CTT GCC TAT GTG GCT CTT GAC TAT GAG CAG GAG ATG AA- ACT GCC AGG AGC AGC

+240 Ser Ser Ile Lys Lys Asn Tyr Glu Leu Pro Asp Gly Gln Asp Ile Ile Ile Gly Ala Glu Arg Phe Arg Cys Pro Glu Val Leu Phe Gln 1169
 TCC TCC ATC AAA AAG AAC TAT GAA TTG CCT GAT GGA CAA GAT ATT ATC ATT GGT GCT GAG AGG TTC CGT TGC CCA GAG GTC CTC TTC CAG

+270 Pro Ser Met Ile Gly Met Glu Ala Ala Gly Ile His Glu Thr Thr Tyr Asn Ser Phe Met Lys Cys Asp Val Asp Ile Gly Lys Asp Leu 1260
 CCA TCT ATG ATT GGT ATG GAA GCT GCA GGTAAATC CAT GAG ACT ACC TAC AAC TCC TTT ATG AAG TGT GAT GAT ATC GGG AAG GAC CTC

+300 Tyr Gly Ile Leu Val Leu Ser Gly Gly Ser Thr Met Phe Pro Gly Ile Val Asp Arg Met Ser Lys Glu Ile Thr Ala Leu Ala Pro Asn 1350
 TAT GGA ATA CTT GTG CTC AGT GGT GGC TCA ACC ATT TTC CCT GGT ATT GTT GAT GAT AGC AAG GAA ATC ACT GCC TTG GCT CCC AAC

+330 Asn Met Lys Ile Lys Val Val Ala Pro Pro Glu Gly Lys Tyr Ser Val Leu Ile Gly Gly Ser Ile Leu Ala Ser Val Ser Thr Phe Gln 1440
 AAC ATG AAG ATC AAG GTT GTT GCT CCA CCG GAG GGA AAG TAC ACT GTC TTG ATT GGA GGA TCC ATC CTA GCG TCA GTT AGC ACC TTC CAG

+360 Gln Met Trp Ile Ser Lys Gly Glu Tyr Asp Glu Ser Gly Pro Ser Ile Phe His Arg Lys Cys Phe ... 1536
 CAG ATG TGG ATT TCA AAG GGC GAG TAC GAT GAG TCT GGT CCA TCC ATT TTC CAC AGG AAG TGT TTC TAA GTGAAGTCTGCTGCTTAGGCAAG

TTCTTTGAGATGGGGCTATGAAGAGGGCCGATTAATAACATTGTACTGCTGACTATTTTCATGTTTCCAGTGCAGCCCTTTTGTCTGGAAGTATCCCTTCTTGAAGTCTGAACTCT 1655
 TTAGGTTGTTGTTAGTTGAGGAGAGTGATTTGTTGTTGTTTATTATTTTGTGSGGTTTTATCTTTTCATTTCCCTTTGTTATTCATTTCTGTTAGTACTTCTXTTTTAACTCTGGGAAC 1774
 AAATGTGTCGATATTTATTGAAAAAATTAGACATTTGCGAGAGCATCCATGATGACTGGAATAGTGTAAAGGTTGGAGGAAAAAGGATGCCCTTTTCACAGATCTTTTCTGCTTG 1893
 AAACCTAAACAAGTATGAGAGATATATTAGCCTGATTATGTAAGGGAAGGAGAGAGCAGTGCTTATTATTCGGAAATAACCTTCAACATTGGATGTAATTCATAATTGACAAAG 2012
 TTAGAGGTTTTCACAGCAACACAGAGATTGCCACTAACATAGAAAGTCCAGATTAGTGGACGTAACATCCCTTTT 2087

Fig. 2 The nucleotide sequence of the gene and its putative protein. Numbers above amino-acid sequence, residue position in the mature protein. Note the two stop codons at positions 89 and 209; numbers at right are nucleotide positions. Note that the five single nucleotide deletions found in this gene are represented by dashes (—) to keep the reading frame in phase. Triangles, the three conserved intron positions, found in all other plant actin genes so far studied. Underline, poly(A) tract; arrows, the 18-bp direct repeats flanking this processed pseudogene (with one mutational difference between them).

presence of multiple frame-shift mutations and two stop codons in its coding sequence. One mutational difference exists between the two 18-bp direct repeats, which may have arisen since the formation of the pseudogene.

Comparison of the sequence of this processed pseudogene with those of five other potato actin genes (G.D., unpublished results) failed to identify the functional gene responsible for its production (results not shown).

The analysis of several mouse, rat and human actin genes has revealed that a large number of the cytoplasmic actin genes in these mammalian species are processed pseudogenes¹¹⁻¹³. However, no processed pseudogenes have yet been found of mammalian muscle actin genes. This further supports the notion that only genes which are expressed in the germ line and which are reverse transcribed and reintegrated back into the germline genome would accumulate through successive generations.

Although all plant actin genes are cytoplasmic actin genes, they probably do not need, as in animals, to be specifically transcribed in meiocytes in order to be found as processed pseudogenes. Reverse transcription and reintegration might

occur in the somatic cells which ultimately differentiate into gametic mother cells. As such, plant processed pseudogenes might exist for any genes which are expressed in the apical meristem. We have no direct evidence as yet whether the actin pseudogene was formed initially in the soma or the meiocytes. Plant cytoplasmic actin genes might be expressed in both somatic and sexual cells, as is the case in animals.

Our findings show that processed pseudogenes are not restricted to animals. The absence of an independent and restricted germ line in plants may allow, given the presence of reverse transcriptase, several genes that are expressed in somatic cells to be reverse transcribed, integrated back into the genome and transmitted genetically.

We thank M. Thangavelu and R. B. Flavell for the EMBL4 phages containing potato actin sequences and P. Oliver for providing us with *recF*⁻ *Escherichia coli* strains. G.D. was supported by a grant from NSERC (Canada) and F.C.A.R. The sequence data in this publication have been submitted to the EMBL/GenBank Libraries under the accession number Y00279.

Received 16 April; accepted 12 June 1987.

- Vanin, E. F. *Biochem. biophys. Acta* **782**, 231-241 (1984).
- Wilde, C. D. *CRC crit. Rev. Biochem.* **19**, 323-352 (1986).
- Weiner, A. M., Deininger, P. L. & Efstratiadis, A. *Rev. Biochem.* **55**, 631-661 (1986).
- Mount, S. M. & Rubin, G. M. *Molec. cell. Biol.* **5**, 1630-1638 (1985).
- Hull, R. & Covey, S. N. *Trends biochem. Sci.* **8**, 119-121 (1983).
- Pfeiffer, P. & Hohn, T. *Cell* **33**, 781-789 (1983).
- Guilley, H., Richards, K. E. & Jonard, G. *EMBO J.* **2**, 277-282 (1983).
- Shepherd, N. S. *et al. Nature* **307**, 185-187 (1984).

- Harris, N. & Flavell, R. B. *Heredity* **57**, 276-277 (1986).
- Johns, M. A., Mottinger, V. & Freeling, M. *EMBO J.* **4**, 1093-1101 (1984).
- Moss, M. & Gallwitz, D. *Nucleic Acids Res.* **10**, 7843-7849 (1982).
- Moss, M. & Gallwitz, D. *EMBO J.* **2**, 757-761 (1983).
- Tokunaga, K., Yoda, K. & Sakiyama, S. *Nucleic Acids Res.* **13**, 3031-3042 (1985).
- Dente, L., Cesareni, G. & Cortese, R. *Nucleic Acids Res.* **11**, 1645-1655 (1983).
- Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E. & Surrey, S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4298-4302 (1982).
- Sanger, F., Nicklen, S. & Coulson, A. R. *Proc. natn. Acad. Sci. USA* **74**, 5463-5467 (1977).

What's new on the laboratory scene

This week's New on the Market feature ends with a glimpse of several of the products on display at next week's VIIth International Congress of Virology meeting in Edmonton, Alberta, Canada.

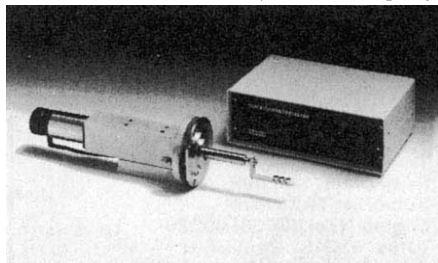
RECENT evidence suggesting that health-care workers can contract the AIDS (acquired immune deficiency syndrome) virus through exposure to blood and body fluids has caused an increase in the care with which such spills are handled. Now, Surgikos Ltd, a Johnson & Johnson company, has come up with a "first-aid kit" approach to **cleaning up body fluid spills** (Reader Service No. 100). The PRESEPT



The portable PRESEPT biohazard disinfection system from Surgikos.

Biohazard Disinfection Kit contains disposable apron and gloves, hypochlorite granules, hypochlorite tablets for preparing a rinsing solution, a plastic scoop and spatula, disposable wipes, and a self-sealing biohazard bag. Surgikos designed the kit for use anywhere where emergency blood spills occur often, such as hospitals, ambulances and police cars. Each £37.50 (UK) kit can handle up to four spills, and refills are available.

Cambridge Technology has expanded the capabilities of X ray detection and analysis with its **trace element detector**, for use with an EDX system in conjunction with scanning electron microscopy (SEM) (Reader Service No. 101). The company



Opens new vistas in X-ray detection.

says the instrument allows various selectable foils to be positioned automatically under the electron beam of the SEM by a microprocessor-controlled motor, and that the resulting transmitted X-ray flux has very low background radiation. Sub-surface information can be obtained with the detector, with no electron beam damage and no surface metallization requirement. Cambridge Technology says the £4,000 (UK) trace element detector has a sensitivity greater than 10 p.p.m.

World Precision Instruments has two new hand-held, battery-operated **pressure monitors** that feature a voltage output connection suitable for chart recorders, oscilloscopes or computers (Reader Service No. 102). The two versions measure pressures in the ranges ± 15 p.s.i., and ± 100 p.s.i. The 100 p.s.i. model also measures pressures in units of



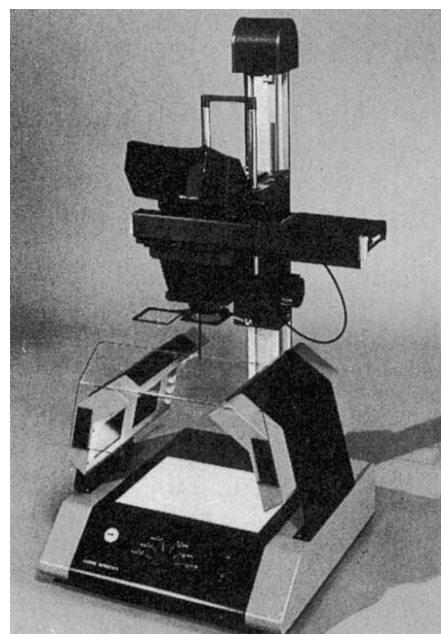
A convenient pressure monitor for a variety of applications.

kPa, and the 15 p.s.i. model also monitors pressures in mm Hg. The monitors are available powered by a 9V alkaline battery for \$395 (US), or with a rechargeable nickel/cadmium battery for \$450 (US), including a 120V battery charger. Barbed pneumatic connectors accept $\frac{1}{8}$ or $\frac{3}{16}$ inch i.d. soft tubing, and four feet of $\frac{1}{8}$ inch i.d. vinyl tubing is included. The monitors' eight ounce weight and $3.1 \times 5.9 \times 1.5$ inch dimensions makes them convenient for a variety of applications.

From Mirage Publishing comes a **map of the atmosphere** that contains a wealth of information on the sea of air that circulates above and around us (Reader

Service No. 103). The 140×85 cm colour map contains information on global atmospheric influences, meteorological motion, severe weather, clouds and precipitation, and meteorological optics (rainbows, halos and mirages). Pollution meteorology, with examples of how man can adversely affect the atmosphere, is also outlined. Information of interest to aviators, including squall lines, gust fronts, wind shear, airborne radar systems and ice accretion is also included. The maps cost \$17.50 (US) each.

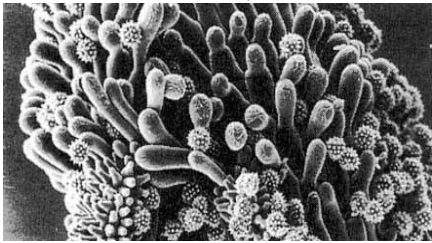
For the **photodocumentation** of thin-layer chromatograms and electrophoretograms, Camag offers its Reprostar/Transilluminator II (Reader Service No. 104). The Reprostar consists of a Polaroid MP-4 multipurpose reflex camera and repro column, side reflectors for direct lighting, and an instrument base containing lamps for transmitted light. The side reflectors come with five 8W lamps each: one short-wave UV (254 nm), two long-wave UV (366 nm) and two white light tubes. The instrument base can include four white tubes and four medium-wave UV (302 nm) lamps, or eight UV lamps. Camag's instrument can be set for short-wave UV direct light,



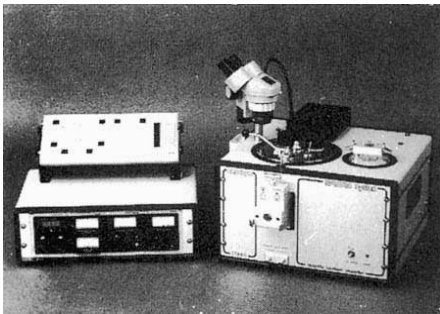
For documenting electrophoretograms.

long-wave UV direct light, direct white light, direct and transmitted white light, or UV 302 nm transmitted light. It comes equipped with a removable UV protection shield.

Low-temperature scanning electron microscopy allows the observation and microanalysis of biological specimens under conditions closely related to their



Frozen hydrated stigmatal surface of alyssum. natural state. Emscope Laboratories Ltd has a **cryopreparation system** to allow researchers to take advantage of this technique (*Reader Service No. 105*). The SP2000A system performs all specimen preparation procedures in an apparatus separate from the SEM, and consists of a macro fracture knife, a macro-manipulator, and stages for sputter coating,



The SP2000A and its system controller. carbon evaporation, metal evaporation and etching. Emscope says the system can be used with all makes of SEM without modification work.

Reading, matching, and compiling data bases of two-dimensional electrophoretograms can be a tedious and error-prone task. Bio Image has introduced a new **machine vision instrument**, the Visage 110, that frees researchers and their technicians from spot-watching (*Reader Service No. 106*). The \$55,000 (US)



A new wave in gel matching technology.

Visage 110 is driven by a Sun Microsystems CPU that has a clock speed of 16.67 MHz, and handles two million instructions per second. The system has a random memory of four megabytes, and a disk storage that can hold 71 megabytes, with a 60 megabyte streaming tape cartridge backup. The Visage 110 combines the image monitor and terminal CRT on one screen, eliminating the multiple terminals of other image analysis systems. Bio Image says the system can scan an entire gel in about one second. The Visage 110 can be purchased with optional \$25,000 (US) Automatic Gel Matcher software, that is capable of matching protein spot patterns as they occur on a series of two-dimensional electrophoretograms.

Virology addendum

The American Type Culture Collection (ATCC) will have several of its **catalogues** on display at the VIIth International Congress of Virology meeting in booth 83 (*Reader Service No. 107*). ATCC's 193-page 1986 catalogue of Animal and Plant Viruses and Antisera describes over 2,000 strains. Its 60-page catalogue of recombinant DNA materials includes vectors, probes, cloned genes, cloned viruses, oncogenes, and their hosts. Also in the ATCC booth will be samples of the Quick Reference Guide for Virus Isolation and Identification. The Guide is a wheel, published by the Centers for Disease Control, that helps virologists identify unknown viruses by matching their characteristics to those of 20 known categories of virus. On the back of the wheel, tips for isolating viruses are spelled out. ATCC is now selling the Guide for \$2.50 (US).

The bacteriophage lambda and its derivative are commonly used cloning vectors for the generation of genomic and cDNA libraries. Bio-Rad's **in vitro packaging kit** uses an infective phage particle system, and contains two separate extracts produced from induced lysogens of two different mutants of lambda (*Reader Service No. 108*). Neither can form a phage particle alone, but when mixed with vector DNA, they can complement each other and assemble particles containing the input DNA. Bio-Rad says up to 1 µg of DNA can be packaged per reaction, and that the reaction is essentially complete after two hours at room temperature. The \$140 (US) kit contains six sets of extracts, *Escherichia coli* LE392 and lambda DNA for testing the system. Bio-Rad will have a display in booth 33.

Becton-Dickinson, in booth 38, has a new monoclonal antibody reagent against the **human T-cell antigen receptor** (*Reader Service No. 109*). Anti-TCR-1 is produced by the WT 31 hybridoma and reacts with a framework determinant of the alpha/beta T-cell antigen receptor

heterodimer, a component of the CD3/Ti cell antigen receptor complex. The antibody is offered in purified form and as biotin and FITC conjugates. Becton-Dickinson says applications currently include the detection and enumeration of cells bearing the T-cell antigen receptor, the identification and isolation of CD⁺/TCR-1⁺ T cells by two-colour immunofluorescence, mitogenic studies of T lymphocytes and characterization of leukaemias and lymphomas.

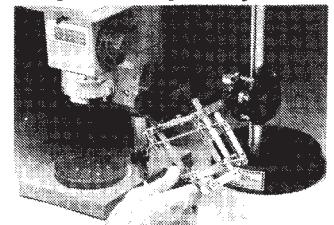
In booth 73, Boehringer Mannheim will have information on its new **pUC cloning kit** (*Reader Service No. 110*). The kit can be used to clone DNA into pUC 18 or 19 vectors. Each kit comes with *E. coli* strains, vectors, T4 DNA ligase, X-gal, IPTG, buffers and protocol. Control DNA — pUC 18 cut with *EcoRI* — is also provided. Boehringer Mannheim guarantees that using the control DNA, its kit will yield 50 per cent as many transformed colonies as obtained using uncut pUC 18, and says results as high as 80 per cent are often observed.

Pharmacia's **power supply** for the agarose gel electrophoresis of nucleic acids will be on display in booths 24/25 (*Reader Service No. 111*). The EPS 250/200 power supply has a 250 V, 200 mA capacity and dual output jacks, that Pharmacia says allows two submarine-style gels to be run simultaneously. It has an audible overcurrent alarm to alert the operator if the current exceeds 200 mA, and carries a \$435 (Cdn) price tag. □

These notes are compiled by Carol Ezzell from information provided by the manufacturers. To obtain further details about these products, use the reader service card bound inside the journal. Prices quoted are sometimes nominal, and apply only within the country indicated.

ADVERTISEMENT

SINGER INSTRUMENTS Xenopus oocyte injection?



SINGER MKI MICROMANIPULATOR

We also offer a Microsyringe Kit.

Write to:

Singer Instrument Co. Ltd.

Treborough Lodge

Roadwater, Watchet

Somerset TA23 0QL

England. Tel. (0984) 40226

Telex: 337492 Comcab G

Reader Service No.50